Hybrid divide-and-conquer approach for tree search algorithms : possibilities and limitations

Mathys Rennela[1], Sebastiaan Brand[2], Alfons Laarman[2], and Vedran Dunjko[2]

[1]Laboratoire de Physique de l'Ecole Normale Supérieure, Inria, CNRS, ENS-PSL, Mines-Paristech, Sorbonne Université, PSL Research University, Paris, France

[2]LIACS, Leiden University, Leiden, The Netherlands

One of the challenges of quantum computers in the near- and mid- term is the limited number of qubits we can use for computations. Finding methods that achieve useful quantum improvements under size limitations is thus a key question in the field. In this vein, it was recently shown that a hybrid classical-quantum method can help provide polynomial speed-ups to classical divide-and-conquer algorithms, even when only given access to a quantum computer much smaller than the problem itself. In this work, we study the hybrid divide-and-conquer method in the context of tree search algorithms, and extend it by including quantum backtracking, which allows better results than previous Grover-based methods. Further, we provide general criteria for threshold-free polynomial speed-ups in the tree search context, and provide a number of examples where polynomial speed ups, using arbitrarily smaller quantum computers, can be obtained. We provide conditions for speedups for the well known algorithm of DPLL, and we prove threshold-free speed-ups for the PPSZ algorithm (the core of the fastest exact Boolean satisfiability solver) for well-behaved classes of formulas. We also provide a simple example where speed-ups can be obtained in an algorithm-independent fashion, under certain well-studied complexity-theoretical assumptions. Finally, we briefly discuss the fundamental limitations of hybrid methods in providing speed-ups for larger problems.

Mathys Rennela: mathys.rennela@inria.fr
Sebastiaan Brand: s.o.brand@liacs.leidenuniv.nl
Alfons Laarman: a.w.laarman@liacs.leidenuniv.nl
Vedran Dunjko: v.dunjko@liacs.leidenuniv.nl

1 Introduction

Years of progress in experimental quantum physics have now brought us to the verge of real-world quantum computers. These devices will, however, for the near term remain quite limited in a number of ways, including fidelities, architectures, decoherence times, and, total qubit numbers. Each of the constraints places challenges on the quantum algorithm designer. Specifically the limitation on qubit count – which is the focus of this work – motivates the search for space-efficient quantum algorithms, and the development of new methods which allow us to beneficially apply smaller devices.

Recent works introduced an approach to extend the applicability of smaller devices by proposing a hybrid divide-and-conquer scheme [1, 2]. This method exploits the prespecified sub-division of problems in such algorithms, and delegates the work to the quantum machine when the instances become small enough. This regular structure also allowed for analytic expressions for the asymptotic run-times of hybrid algorithms.

The hybrid divide-and-conquer method was applied to two cases of divide and conquer algorithms, that of derandomized Schöning's algorithm for solving Boolean satisfiability [1], and to the problem of finding Hamilton cycles on cubic graphs [2]. These schemes achieved asymptotic polynomial speedups given a quantum computer of size m, where m is a fraction of the instance size n, i.e., $m = \kappa n$. Interestingly, these speed-ups are obtainable for all fractions κ, i.e. the im-

provements are *threshold free*.

The space efficiency of the quantum subroutines was identified as a key criterion for determining whether threshold-free speed-ups are possible, as one may expect.

In these works, the quantum algorithmic backbone was Grover's search, which is space-frugal, but known to be sub-optimal for the cases when the underlying search spaces, the *search trees*, are not complete nor uniform. To obtain speedup in these cases, more involved quantum search techniques, namely quantum backtracking [3], need to be employed. However, until this work, it was not clear how the space demands of quantum backtracking would influence the applicability of the hybrid approach.

Here, we resolve this issue, and investigate the generalizations of the hybrid divide-and-conquer scheme from the perspective of algorithms which reduce to tree search, in particular, backtracking algorithms. Our approach is then applied to the two of the arguably best known exact algorithms for Boolean satisfiability: the algorithm of Davis-Putnam-Logemann-Loveland (DPLL) algorithm [4] (which is still the backbone of many heuristic real-world SAT solvers) and the Paturi-Pudlák-Saks-Zane (PPSZ) algorithm [5] (which is the backbone of the best-known exact SAT solver).

The main contributions of this work are summarized as follows:

- We analyze the hybrid divide-and-conquer scheme from the perspective of search trees and provide very general criteria which can ensure polynomial-time speed-ups over classical algorithms (Section 3). We also consider the limitations of the scheme in the context of online classical algorithms, which terminate as soon as a result is found.

- We demonstrate that quantum backtracking methods can be employed in a hybrid scheme. This implies an improvement over the previous hybrid algorithm for Hamilton cycles on degree 3 graphs. While the performance of Grover's search and the quantum backtracking algorithm have been compared in the context of the k-SAT problem, with hardware limitations in mind [6], this is the first time that the hybrid method and quantum backtracking is combined.

- We exhibit the first, and very simple example of an algorithm-independent provable hybrid speed-up with quantum backtracking, under well-studied complexity-theoretic assumptions.

- We study a number of settings from the search tree structure and provide all algorithmic elements required for space-efficient quantum hybrid enhancements of DPLL and PPSZ; specially for the case of PPSZ tree search, we demonstrate, under some assumptions on the variable order, a threshold-free hybrid speed-up for a class of formulas, including settings where our methods likely beat not just PPSZ tree search but any classical algorithm.

- We discuss of the fundamental limitations of our and related hybrid methods.

To achieve the above results, we provide space and time-efficient quantum versions of various subroutines specific to these algorithms, but also of routines which may be of independent interest. This includes a simple yet exponentially more space efficient implementation of the phase estimation step in quantum backtracking, over the direct implementation [7].

The structure of the paper is as follows. The background material is discussed in section 2. This section lays the groundwork for hybrid algorithms from a search tree perspective, by elaborating on how backtracking defines those trees. Section 3 then introduces a tree decomposition which defines our hybrid strategy, i.e. from what points in the search tree will we start using the quantum computer, and analyzes their impact. In Section 3.2, we discuss sufficient criteria for attaining speedups with both Grover-based search and quantum backtracking over the original classical algorithms, including online algorithms. Section 4 provides concrete examples of algorithms and problem classes where threshold-free, and algorithm independent speed-ups can be obtained. Finally, in Section 5, we discuss the potential and limitations of hybrid approaches for the DPLL algorithm, also in the more practical case when all run-times are restricted to polynomial. This section also briefly addresses the question of the limits of possible speed-ups in any hybrid setting. The appendix collects all

our more technical results, and some of the background frameworks.

2 Background

This section introduces SAT and a general backtracking framework, which is instantiated for two exact algorithms: DPLL and PPSZ. While DPLL performs better in practice on many instances, PPSZ-based algorithms provide the best worst-case runtime guarantee at the time of writing [8]. The section ends with an explanation of the hybrid divide-and-conquer method.

2.1 Satisfiability

A Boolean formula F over n variables corresponds to a function $F : \{0,1\}^n \to \{0,1\}$ in the natural way. The Boolean satisfiability problem (SAT) is the constraint satisfaction problem of determining whether a given Boolean formula F in conjunctive normal form (CNF) has a satisfying assignment, i.e. a bit string $y \in \{0,1\}^n$ such that $F(y) = 1$. A CNF formula is a conjunction (logical 'and') of disjunctions (logical 'or') over variables or their negations (jointly called literals). For ease of manipulation, we view CNF formulas as a set of clauses (a conjunction), where each clause is a set of k literals (a disjunction), with positive or negative polarity (i.e. a Boolean atom x, or its negation $\bar{x}$). In the k-SAT problem, the formula F is a k-CNF formula, meaning that all the clauses have k literals.[1]

It is well known that solving SAT for 3-CNF formulas (3-SAT) is an NP-complete problem. As a canonical problem, it is highly relevant both in computer science [9] and outside, e.g. finding ground energies of classical systems can often reduced to SAT (see e.g. [10]).

Since it is a disjunction, a clause $C \in F$ evaluates to true (1) on a (partial) assignment $\vec{x}$, if at least one of the literals in C attains the value true (1) according to $\vec{x}$. A formula F evaluates to true on an assignment $\vec{x}$, i.e. $F_{\vec{x}} = 1$, if all its clauses evaluate to true on this partial assignment. Note that to determine the formula evaluates to true on some (partial) assignment, the assignment has to fix of at least one variable per clause of F, i.e. the assignment should be linear in the number of variables. On the other hand, a formula can evaluate to false (0) on very sparse partial assignments: it suffices that the given partial assignment renders any of the clauses false by setting k variables, i.e. a constant amount. In such a case, we say that the partial assignment *establishes a contradiction*. This is the equivalent of saying that the constraint formula is *inconsistent*.

Given a partial assignment $\vec{x} \in \{0,1,*\}^n$, we denote with $F_{|\vec{x}}$ the subfunction of F restricted to that assignment. This subfunction can also be obtained by setting the variables in the formula F to the values specified in the partial assignment. For CNF formulas this means the following: for an assigned variable x_j, for every clause of F where x_j appears as a literal (of some polarity), and the setting of x_j renders the corresponding literal true, that clause is dropped in $F_{|\vec{x}}$. For every clause of F where x_j appears as a literal (of some polarity), and the setting of x_j renders the corresponding literal false, that literal is removed from the corresponding clause. Finally, an assignment $\vec{x}$ that establishes a contradiction introduces an *empty clause*, i.e. $\emptyset \in F_{|\vec{x}}$, whereas a satisfying assignment $\vec{y}$ yields an *empty formula*, i.e. $F_{|\vec{y}} = \emptyset$. We will call such formulas, where some variables have been fixed ($F_{|\vec{x}}$), *restricted formulas*, and say that $F_{|\vec{x}}$ is a restriction of F by the partial assignment $\vec{x}$. We will also use a similar notation for setting literals to true. Given a literal $l \in \{x_i, \bar{x}_i\}$ we write $F_{|l}$ for the restricted formula given by F with the value of the variable x_i set such as to render l true.

Additionally, we say a formula G is semantically entailed by a formula F if all satisfying assignments (models) of F also satisfy G, denoted $F \models G$. In general G can be any formula over a subset of variables of F, although we only consider cases where G consists of a single literal $l \in \{x_i, \bar{x}_i\}$.

Finally, *resolution* is a logical inference rule used for satisfiability proving [11]. Basic resolution rules takes two clauses $(x \vee A)$ and $(\bar{x} \vee B)$, where A, B are clauses, and derives a clause $(A \vee B)$, since both clauses are satisfiable if and only the inferred clause is. The clausal inference rule is refutation complete, meaning that a complete, recursive search over all inferred clauses will find a refutation if the original formula, i.e. a

[1] Without loss of generality, we will allow that the formula has clauses with fewer literals, but assume that at least one clause has k literals, and no clauses have more.

Algorithm 1 The DPLL algorithm

```
1: function DPLL(F: CNF)
2:     if F = ∅ then return 1
3:     if ∅ ∈ F then return 0
4:     x ← next var acc. to branching heuristic
5:     if F ⊨_h (x = c) with c ∈ {0,1} then
6:         return DPLL(F_{|x=c})
7:     else
8:         return DPLL(F_{|x}) ∨ DPLL(F_{|x̄})
```

set of clauses, is unsatisfiable. In *unit resolution*, we have $A = \emptyset$ (or $B = \emptyset$) in the above, i.e., one of the input clauses is a *unit clause* (x) (or $(\bar{x})$).

2.2 The DPLL algorithm family

The algorithm of Davis Putnam Logemann Loveland (DPLL) is a backtracking algorithm which recursively explores the possible assignments for a given formula, applying (incomplete) reduction rules along the way to prune the search space [4]. Originally designed to resolve the memory intensity of solvers based only on resolution (the DP algorithm uses a complete resolution system [11]), this backtracking-based algorithm can now be found in some of the most competitive SAT solvers [12]. DPLL was also generalized to support various theories, including linear integer arithmetic and uninterpreted functions [13], and is using consequently in many combinatorial domains [9] and automated theorem proving [14].

DPLL branches on assignments to individual variables x and $\bar{x}$. It uses *reduction* rules based on resolution to prune the resulting search tree. In order to prune a branch, a reduction rule $\mathcal{R}$ (efficiently) *under-estimates* whether a literal l is entailed by a formula F, i.e.,

$$F \models_{\mathcal{R}} l \implies F \models l.$$

We consider two rules:

- Unit resolution: if there exists a unit clause $\{x\}$ (or $\{\bar{x}\}$), then set value x to true (false).
- Pure literal: if variable x only appears positively (negatively), then set x to true (false).

Algorithm 1 corresponds to the DPLL algorithm in its classical form. It recursively assigns truth values to the variables of F in a heuristically chosen order (see Line 4), while eagerly simplifying the formula (see $F_{|x}, F_{|\bar{x}}$ at Line 6,8). Branches are possibly pruned with the reduction rule at Line 6. When the formula becomes (un)satisfiable, the recursion backtracks (Line 2,3). Note that a short circuiting or at Line 8 will cause the algorithm to terminate early when the formula is found to be satisfiable.

The algorithm explores a subtree of the full binary search tree. We can consider each node of the tree as uniquely labeled with a (restricted) formula or a partial assignment, as explained in more detail in section 3. We call the nodes where pruning happens, the *forced nodes* and the others *guessed nodes.*

A key element of the (heuristic) efficiency of backtracking algorithms, is its ability to simplify subproblems as early as possible in order to prune the tree search. This is what DPLL achieves through its branching heuristic, which determines the order in which variables should be considered. Notice that the order may change in the different branches of the tree, for example, because different unit clauses appear under different assignments in the concrete heuristic discussed above. On the other hand, one can of course always implement a static branching heuristic that simply selects variables according to a predetermined order, as we do in the next section for PPSZ.

2.3 The PPSZ algorithm family

The best exact algorithms for the (unique) k-SAT problem have for many years been based on the PPSZ (Paturi, Pudlák, Saks, and Zane [5]) algorithm. The PPSZ algorithm is a Monte Carlo algorithm, i.e. it returns the correct answer with high probability using randomization, although a derandomized version also exists [15]. Like DPLL, PPSZ uses heuristics to force or guess variables. The upper bound on the runtime is derived from the probability that a variable is guessed, as the expected number of guesses dictates the size of the search space. We first explain the intuition based on a simple version that achieves a runtime of $O^*(2^{(1-1/k+\varepsilon)n})$, for $\varepsilon > 0$.[2]

[2]In the remainder of the text, we will be using the standard notation for "lazy" scalings: with the superscript $*$, e.g. O^*, we denote scalings which ignore only polynomi-

Algorithm 2 dncPPSZ$_s(F,\pi)$

if $F = \emptyset$ **then return** 1
if $\emptyset \in F$ **or** $n - |\pi| > (\gamma_k + \varepsilon)n$ **then return** 0
$x,\ \pi \leftarrow \pi[0],\ \pi[1\ldots]$ ▷ first var in π, postfix
if $F \models_s (x = c)$ **with** $c \in \{0,1\}$ **then**
    return dncPPSZ$_s(F_{|x=c}, \pi)$
else
    return dncPPSZ$_s(F_{|x},\pi) \vee$ dncPPSZ$(F_{|\bar{x}},\pi)$

Assuming there is only a single satisfying assignment $\vec{u}$, it is easy to see that the unit resolution heuristic can force on average $1/k$ variables for a random variable permutation $\pi \in S_n$: In the first place, for every variable x there has to be a clause C which forces the value of x, i.e., where the only true literal under assignment $\vec{u}$ is x or $\bar{x}$, otherwise there would be multiple satisfying assignments. Second, if the variables are assigned according to the order π and π places x last with respect to the other $k-1$ variables of C, then unit resolution will indeed identify the forced variable. As a result, a PPSZ algorithm merely has to search a space proportional to the number of guesses. For a k-SAT problem we let γ_k denote the expected fraction of the n variables which are guessed by PPSZ, such that $\gamma_k n$ is the expected number of guessed variables. The amount of variables which need to be guessed depends on the heuristic h used to check if $F \models_h l$. If the heuristic consists solely of unit resolution,[3] then for unique[4] k-SAT PPSZ needs to make at most $(1 - 1/k)n$ guesses. When unit resolution is replaced with the more general s-implication (discussed below) something equivalent to the original PPSZ algorithm is obtained.

There are currently various versions of PPSZ (see e.g. [17, 18, 19, 8]), but recent analysis [20] shows that the original PPSZ [5] still has the best worst-case runtime of $2^{(\gamma_k+\varepsilon)n}$, where for example $\gamma_3 \approx 0.386229$. The original PPSZ can be seen as using a reduction rule called s-implication during the tree search [18]. A literal $l = x, \bar{x}$ is s-implied, written $F \models_s l$, if there is a sub formula of s clauses, which implies it, i.e. $G \subseteq F$ with $|G| = s$ and $G \models l$. In other words, if all satisfying assignments of G set the variable x to the same value.

The original PPSZ algorithm evaluates random assignments on random variable orders and is not immediately amenable for (quantum) backtracking. Here we introduce a backtracking version of PPSZ that offers the same runtime guarantees as recent PPSZ versions, while allowing for the benefits of backtracking on a quantum computer (i.e., heuristically pruning search). Algorithm 2 shows dncPPSZ, which is similar to DPLL with s-implication as reduction rule. It takes a fixed (random) variable order π and prunes branches after $(\gamma_k + \varepsilon)n$ guesses have been performed. Since the expected number of guesses over all variable orders equals $\gamma_k n$ for s-implication [16], Markov's inequality tells us that, for any $\varepsilon > 0 \in \Theta(1)$, with constant probability we can find the satisfying with fewer than $(\gamma_k + \varepsilon)n$ guesses for a random π. Hence we obtain Proposition 2.1 (see Appendix A for more detail and a proof).

Proposition 2.1. *Executing dncPPSZ a constant number of times, using random variable orders π, is sufficient to decide satisfiability. The run-time of dncPPSZ is upper bounded by the run-time of the standard PPSZ algorithm.*

We highlight that the PPSZ algorithm involves two steps: the "core" of the algorithm which is the *PPSZ tree search* performed by the dncPPSZ algorithm, which takes an ordered formula on input (i.e. the variable ordering is fixed, by e.g. the natural ordering over the indices). And an outer loop that repeats the PPSZ tree search a constant number of times, randomizing the order for each call. This is critical in our subsequent analysis.

2.4 Quantum algorithms for tree search

The extent to which quantum computing can help the exploration of trees, and more generally, graphs is a long-standing question with infrequent but noticeable progress. In the context of backtracking (for concreteness, for SAT problems), up until relatively recently, the best methods involved Grover search [21] over all possible

ally contributing terms (relevant when the main costs scale exponentially), just as tilde, e.g. $\tilde{O}$ highlights we ignore logarithmically contributing terms (when the main costs are polynomial in the relevant parameters).

[3] The PPSZ algorithm where unit resolution is the only heuristic is in fact the PPZ algorithm [16].

[4] This approach can be generalized to a setting with multiple satisfying assignments [16].

satisfying assignments (the leafs in the full search tree). In such an approach, we would introduce a separate register of length d equal to the number of variables and a *search predicate*, implementing the satisfiability criterion based on the input formula. This brute-force approach however yields a redundant search over branches which would be pruned away by the resolution rules, and, in the worst case may force a quantum computer to search an exponentially larger space than the classical algorithm would.

Nonetheless, Grover provides an advantageous strategy whenever the classical search space is larger than $2^{n/2}$, hence this was the method used in previous work on hybrid divide-and-conquer strategies [2, 1]. One example is Schöning's algorithm for SAT, which allows a full quadratic quantum speed-up [22] (but was since surpassed by PPSZ). A particular advantage of Grover's search is that it is frugal regarding time and space: beyond what is needed to implement the oracle (the search predicate), it requires at most one ancillary qubit, and very few other gates.

More recently, Montanaro [3], Ambainis & Kokainis [23], and Jarret & Wan [24] have given quantum-walk based algorithms which allow us to achieve an essentially full quadratic speed-ups in the number of queries to the tree (when trees are exponentially sized). We will refer to the underlying method as the *quantum backtracking method*.

The quantum backtracking method can be applied whenever we have access to local algorithms which specify the children of a given vertex, and whenever we can implement the search predicate, indicating the satisfiability of a (partial) assignment. At the heart of the routine is the construction of a Szegedy-style walk operator W over the bipartite graph specified by the even and odd depths of the search tree (details provided in Appendix B.2); Montanaro shows that the spectral gap of W reveals whether the underlying graph contains a marked element (a satisfying assignment, as defined by the search predicate P). The difference in the eigenphases of the two cases dictates the overall run-time, as they are detected by a *quantum phase estimation* (QPE) overarching routine. The overall algorithm calls the operator W no more than $O(\sqrt{Tn}\log(1/\delta))$ times, for a correct evaluation given a tree of size T, over n variables (depth), except with probability δ. This is essentially a full quadratic improvement when T is exponential (even superpolynomial will do) in n. For the algorithm to work, one assumes that the tree size is known in advance, or at least, a good upper bound is known.

The original paper on quantum backtracking implements DPLL with the unit rule [3], and led to a body of work focused on improvements [23, 24] and applications [7, 25, 23, 6].

Tree search implementation. The implementation details are important for the discussion of the efficiency of later approaches, so we provide a framework for the implementation of quantum backtracking. If the functions specified in the framework are implemented (reversibly), then quantum backtracking can be implemented with almost no space overhead, as we provide a space efficient implementation of QPE (see section B.2). The runtime overhead is multiplicative, so any polynomial implementation of the framework will do since we focus on obtaining a reduction in the exponent in the hybrid method.

The trees $\mathcal{T}$ that we consider are characterized by the backtracking functions specifying the local tree structure for some input formula F, as described in Section 2. The nodes v of the tree contain the information to represent the restriction $F_{|\vec{x}}$ for some partial variable assignment $\vec{x}$. In the classical approach (DPLL) each node is represented by the restricted formula itself (as a clause database [9]).

Here, we highlight a duality between partial assignments $\vec{x}$ and corresponding restricted formulas $F_{|\vec{x}}$ – instead of defining the tree in terms of partial assignments, one can define it in terms of restricted formulas, and we will often use this duality as storing partial assignments for a fixed formula requires less memory than storing a whole restricted formula. This duality is also important as it allows us to see backtracking algorithms as divide-and-conquer algorithms: each vertex in a search tree – the restricted formula $F_{|\vec{x}}$ – corresponds to a restriction of the initial problem, and children in turn correspond to smaller instances where one additional variable is restricted. Consequently backtracking algorithms can often be neatly rewritten in a recursive form.

To implement the walk operator for quantum backtracking, it suffices to be able to implement a unitary operator which given a vertex v, produces

the children of v and one that decides whether the partial assignment corresponding to node v makes the formula true. For this, we specify the functions $ch1, ch2, chNo, P$ below. Appendix C shows how the walk operator can be implemented based on these functions.

- a function $ch2(v, b)$, which takes on input a vertex v, and returns one of the children, specified by the bit b;
- a function $ch1(v)$ which returns the single child if v is forced;
- a function $chNo(v)$, which returns whether v has zero, one or two children, and
- a search predicate P which identifies leafs and solutions, i.e., $P(\vec{x}) = b$ if $F_{|\vec{x}} = b$ with $b \in \{0, 1\}$, and $P(\vec{x}) = \perp$ otherwise.

As in many cases, the classical, non-reversible implementations are efficient, the runtime of the reversible version is often not a problem, as long as it stays polynomial (although even sub exponential will suffice). However, in the context of hybrid methods, the space complexity of these implementations becomes vital. Standard approaches efficiently implement a data structure which can represent the last vertex visited, and update this structure with the next vertex reversibly. The reversibility prevents deleting information efficiently, so the simplest solution is to store the entire branching sequence, reconstructing the state from this succinct representation. Indeed, the technical core of the papers [2, 1], and parts of the remainder of this paper, focus on low space algorithms solving this problem.

Online algorithms. Note, the quantum algorithm for quantum backtracking always performs $O(\sqrt{Tn}\log(1/\delta))$ queries no matter which vertex/leaf is satisfying, if any. In contrast, classical backtracking is an *online algorithm*, meaning that it can terminate the search early when a satisfying assignment is found. This naturally depends on the traversal order of the tree, which is also specified by the algorithm (and perhaps the random sequence specifying any random choices), and thus the maximal speed-ups are only achieved for the classical worst cases when either no vertices are satisfying, or when the very last leaf to be traversed is the solution.

In [23], the authors provide an extension to quantum backtracking, which allows one to estimate the size of the search tree. Moreover, it can also estimate the size of a search space corresponding to a partial traversal of a given classical backtracking algorithm, according to a certain order.

With this, it is possible to achieve quadratic speed ups not in terms of the overall tree, but rather the search tree limited to those vertices that a classical algorithm would traverse, before finding a satisfying assignment. In this case, we have a near-full quadratic improvement, whenever this effective search tree, which depends on the search order, is large enough (superpolynomial).

2.4.1 Grover vs quantum backtracking in tree search

One can employ Grover search-based techniques to explore trees of any density and shape. Yet in many cases using Grover can be significantly slower than by employing the quantum backtracking strategy, or even classical search. While the query complexities of classical and quantum backtracking depend on the tree size, the efficiency of quantum search based on Grover is bounded by the maximal number of branches that occur in any path from root to the leaf of the tree. The branching number is also vital in space complexity analyses.

Definition 2.2 (Maximal branching number)**.** *Given a tree $\mathcal{T}$, the maximal branching number $br(\mathcal{T})$ is the maximal number of branchings on any path from root to a leaf. If a sub-tree $\mathcal{T}'$, is specified by its root vertex v, with $br(v)$ we denote $br(\mathcal{T}')$.*

Here is a sketch of how Grover's search would be applied in tree search leading to the query complexity dependence on branching numbers. For the moment, we assume that all the leaves of the tree are at distance n from the root; this can be obtained by attaching sufficiently long line graphs to each leaf occurring earlier. This may cause a blow up of the tree by at most a factor of n, but this will not matter in the cases where trees are (much) lager, i.e. exponential.

Our application of Grover's search over a rooted tree structure follows the following principle. We make use of an "advice" register, which

tells the algorithm which child to take, when two options are possible, i.e. select the b parameter of $ch2$ when $chNo(v) = 2$. Let $v_1, v_2, \ldots, v_n$ be a path from the root $v_1 = r$ to a leaf v_n. For each guessed v_i along the path, i.e. where $chNo(v_i) = 2$, a subsequent unused bit of the advice string determine the choice. Finally, the n^{th} vertex is checked by the search predicate P. For this process to be well defined, we clearly require as many bits in the advice register as the largest number of branches along any path.

We now show that the size of the advice can also be independent from the actual tree size. To do so, we first discuss tree shapes for which Grover exhibits extremal behavior.

Take a "comb" graph for instance, where one edge is added to every node of a line graph, except to the last node. This graph has $2n - 1$ vertices, and $(n - 1)$ branches whereas a full binary tree with 2^n vertices has the same maximal number of branches (this disparity persists even when we "complete" the comb graph by extending the single line to ensure every path is length n). The Grover-based algorithm thus always introduces an effective tree which is exponential in the number of branchings.

Note this need not lead to exponential search times; e.g., in the example of the comb graph, if the leaf-child of the root is satisfying, then half $(2^{n-1} = 2^n/2)$ of the strings in the advice register will result in finding the satisfying assignment,[5] leading to an overall constant query complexity.

More generally, in the case of search with a single satisfying assignment v, the worst-case query complexity using Grover's approach, will always be $O(2^{b/2})$, where b is the number of branches on the path from the root to v, i.e, $b = br(r) - br(v)$, where r is the root of the tree. This is because all the vertices below v will represent solutions, as explained in the previous paragraph.

Comparing Grover's search with backtracking in the sense of query complexity, we find the following: any binary tree of size $O(2^{\gamma n})$ must contain a path from the root to a leaf which has more than γn branches – if the tree contains no more than γn branches along any path, i.e. its branching number is γn, then Grover's method will achieve a run-time of $O^*(2^{\gamma n/2})$ as well.

In other words, a limitation on the tree size – which upper-bounds the classical search run-time – directly limits the quantum backtracking query complexity, whereas it does not (as directly) limit Grover-based search. However, as we will show in the case of PPSZ, when the tree size is estimated on the basis of (γn) maximum branching numbers, this does provide a way to directly connect classical search with Grover-based method.

Unlike backtracking, Grover can also be used as an online algorithm, as discussed earlier in this section. In the majority of this paper, we will be concerned with worst-case times, so the online methods of [23] will not be critical. Although, for practical speed-ups, they certainly are. We reiterate that all the results that we will present, can accommodate these tree size-estimation based methods of [23].

2.5 The hybrid divide-and-conquer method

The hybrid divide-and-conquer method was introduced to investigate to which extent smaller quantum computers can help in solving larger instances of interesting problems. Here the emphasis is placed on problems which are typically tackled by a divide-and-conquer method. This choice is one of convenience. Any method which enables a smaller quantum computer to aid in a computation of a larger problem must somehow reduce the problem to a number of smaller problems. How this can be done is critical for the hybrid algorithm performance.

But in divide-and-conquer strategies, there exists an obvious solution. Divide-and-conquer strategies recursively break down an instance of a problem into smaller sub-instances, so at some point they become suitably small to be run on a quantum device of (almost) any size.[6]

In previous work on the hybrid divide-and-conquer method [2, 1] this approach was used for a de-randomized version of the algorithm of Schöning, and for an algorithm for finding Hamilton cycles in degree-3 graphs. In general, in these works (and the present work) the question of interest is to identify criteria when speed-ups, in the sense of provable asymptotic run-times of the (hybrid) algorithms, are possible.

[5] If it holds that $P(v) = 1$ for node v, then $P(v') = 1$ for all descendants of v' of v.

[6] More precisely, the device must large enough to handle any size instance of a given problem, which is in not a trivial condition.

2.5.1 Quantifying speed-ups of hybrid divide-and-conquer methods

For the quantum part of the computation, the complexity-theoretic analysis of meta-algorithms (more precisely, oracular algorithms), such as Grover and quantum backtracking, predominantly measures *query complexity*. This is the number of calls to the e.g. Grover oracle, or the walk operator, respectively, i.e. the black boxes that implement the predicate detection, and the search tree structure.

As the present work is only concerned with exponentially sized trees, and sub-exponential time sub-routines, the complexity measure we use for the classical algorithm $\mathcal{A}_C(n)$ is the size of the tree (i.e, the classical query complexity) and for quantum algorithm the query complexity. In the hybrid cases, we will thus measure the totals of classical and quantum query complexities, treated on an equal footing.

We are thus interested in the (provable) relationships between quantities $\mathrm{Time}(\mathcal{A}_C(n))$ describing the run-time of the classical algorithm given instance size n, and $\mathrm{Time}(\mathcal{A}_H(n,\kappa))$,[7] describing the run-time of the hybrid algorithm, having access to a quantum computer of size $m = \kappa n$ with $\kappa = O(1)$.[8] We focus on exact algorithms for NP-hard problems (which typically have exponential runtime).

Definition 2.3 (Hybrid speed-ups)**.**
We say that genuine *speed-ups (i.e. polynomial speed-ups) are possible using the hybrid method, if there exists a hybrid algorithm such that*

$$\mathrm{Time}(\mathcal{A}_H(n,\kappa)) = O(\mathrm{Time}(\mathcal{A}_C(n))^{1-\epsilon_\kappa}), \quad (1)$$

for a constant $\epsilon_\kappa > 0$. If such an ϵ_κ exists for all $\kappa > 0$, then we say the speed-up is threshold-free*.*

What originally sparked the interest in hybrid algorithms, is the fundamental question whether *threshold-free* speed-ups are possible at all. This was answered in the positive in the previous works mentioned above [2, 1].

In the above, we have assumed that the time complexity can be fully characterized in terms of the instance size n. In general, as discussed in Section 2.4.1, the complexities may be precisely established only given access to a number of parameters, such as, search tree size or even less obvious measures like the location of the tree leaf in the search tree.

2.5.2 Limitations of existing hybrid divide-and-conquer methods

The key impediment to threshold-free speed-ups is the space-complexity of the quantum algorithm for the problem. To understand this, assume that with $\kappa' n$ we denote the size of the instance we can solve on a κn sized device (so $\kappa n = \mathrm{Space}(\mathcal{A}_Q(\kappa' n))$, where $\mathrm{Space}(\mathcal{A}_Q(n))$ denotes the space complexity of the quantum algorithm). If the space complexity is super-linear, then κ' itself becomes dependent on n, and in fact, decreasing in n. In other words, as the instance grows, the *effective fraction of the problem* that we can delegate to the quantum device decreases. Then, in the limit, all work is done by the classical device, so no speed-ups are possible (for details see [2, 1]).

In [2], these observations also lead to a characterization of when genuine speed-ups are possible for recursive algorithms, whose run-times are evaluated using standard recurrence relations. Notably, none of the classical algorithms employed backtracking nor early pruning of the trees. These properties ensured that the search space could be expressed as a sufficiently dense trees, ensuring that no matter where we employ the (faster) quantum device, a substantial fraction of the overall work will be done by the quantum device, yielding a genuine speed-up.

Consequently, the main technical research focus in these earlier works were to establish highly space-efficient variants of otherwise simple quantum algorithms (in essence, at most linear in n), which are all based on Grover's algorithm over appropriate search spaces. Both the derandomized algorithm of Schöning [1], and Eppstein's algorithm for Hamilton cycles [2] traverse search trees dense enough that Grover-based search (which is itself space-efficient) over appropriate search spaces, yields a polynomial

[7] As we will discuss later, in the cases of the quantum algorithms we will consider, the relevant notion of instance size may be different from what is relevant for the classical algorithm complexity. However to compare the run times, we will always have to bound both hybrid and classical complexities in terms of same quantities.

[8] Since we consider speed-ups in the sense of asymptotic run-times, considering smaller sizes, e.g. constant sized quantum computers, or log sized quantum computers makes little sense as these are efficiently simulatable.

improvement. (Appendix D describes an improvement over the hybrid solution of [2] based on the new framework developed in Section 3.)

An additional limitation is that any speed-up obtained through this hybrid approach is always relative to a fixed classical algorithm, as established in the previous works. This also implies that even in the event that we obtain a hybrid speed-up for the best algorithm for a given problem, this speed-up, and specially, any chance of a threshold free speed-ups disappear whenever a new faster algorithm is devised. In other words, the speed-ups are algorithm-specific. This will always be true, unless lower bounds are proved for the problem that these classical algorithms attack, in which case we may talk about *algorithm independent* speed-ups.

In the following two sections we study hybrid divide-and-conquer in the context of search trees, and provide a number of new algorithms using the hybrid method.

3 Hybrid divide-and-conquer for backtracking algorithms

In the present work, we consider the hybrid divide-and-conquer method for algorithms which operate by searching over a suitable tree. This framework naturally captures backtracking algorithms, and recursive algorithms as studied in [2]. Our new framework focuses on scenarios in which the search is over unbalanced trees, unlike the Grover-based methods utilized in [2, 1].

This section investigates the structure and properties of hybrid divide and conquer algorithms from the search tree perspective. Most notably, it introduces the concept of tree search decomposition. These considerations influence the design of the new hybrid divide-and-conquer algorithms discussed later.

The outline of this section is as follows. In Subsection 3.1, we discuss how specifically the tree structure influences whether or not (provable) polynomial speed-ups can be obtained on an intuitive level. Then, in Subsection 3.2, we provide a theorem providing a general, albeit not very operative, characterization of when speed-ups are possible. In Subsection 3.3, we connect more closely the properties of the quantum algorithms with the tree structure identifying assumptions which allow for significant simplifications. In Subsection 3.4, quantitative speed-ups are proved for constrained (but still rather generic) cases. These special cases are then exemplified in Section 4, and Section 5 addresses some of the scenarios were these simplifying conditions cannot be met.

3.1 Search tree structure and potential for hybrid speed-ups

In previous works, the importance of the space efficiency of quantum algorithms was put forward as the key factor determining whether asymptotic (threshold-free) hybrid speed-ups can be achieved, as discussed in detail in [2]. Naturally, in the hybrid backtracking setting, we will inherit the same limitations, and some new ones, which we focus on next.

A crucial aspect of our hybrid framework is that we always assume that the size of the quantum computer is κn, where $\kappa \in (0, 1]$ and n is a *natural* problem instance size. Note that the instance size is not unambiguous (as discussed in Section 5). Nonetheless, we assume this is well-defined, and that the classical backtracking algorithm we consider generates search trees of height (at most) n.[9]

Next, to understand how the tree structure may influence the overall algorithm performance, we introduce the search tree decomposition. Consider a search tree $\mathcal{T}$ generated by algorithm $\mathcal{A}$ on some instance of size n, where every vertex in the tree denotes a sub-problem which can (in principle) be delegated to a quantum computer, running the algorithm $\mathcal{A}_Q$ in a hybrid scheme, which we will denote with subscript H. Note, $\mathcal{A}_Q$ can be the quantum backtracking or Grover version of $\mathcal{A}$, or a related algorithm for solving the same problem.

In general, κn does not provide enough space for $\mathcal{A}_Q$ to be run on the entire problem, i.e. on the root of the tree, but it can be run on some of sub-instances represented by vertices deeper in the tree. This of course depends on the space efficiency features of $\mathcal{A}_Q$ – the *effective size* of the quantum computer we have, but we will not focus on this for the moment.

[9] In the next section's examples, with the exception of the enhanced hybrid algorithm for Hamilton cycles on cubic graphs discussed in Appendix D, n specifically designates the number of variables of a given Boolean formula (and not the formula length itself).

We merely assume that whether or not $\mathcal{A}_Q$ can be run on an instance v given an κn-sized device is monotonic with respect to the tree structure; that is, if it can be run on v, it can be run on all descendants of v.[10]

With this in mind, we can identify the collection of J_c *cut-off vertices* $\{c_k\}_{k=1}^{J_c}$. That is, all the vertices that can be run on the quantum device, whose parents cannot be run on the same device due to size considerations. In principle, J_c may be zero for some trees. These vertices correspond to the largest instances in the search tree where we can use the quantum computer.

Now the *search tree decomposition* is characterized by the set of sub-trees $\{\mathcal{T}_j\}_{j=0}^{J_c}$, where the subtree $\mathcal{T}_j$ is the entire sub-tree rooted at the cut-off vertex c_j, and where $\mathcal{T}_0$ denotes the tree rooted at the root of $\mathcal{T}$, whose leaves are parents of the cut-points $\{c_j\}_j$. We refer to $\mathcal{T}_0$ as the "top of the tree", and it is traversed by the classical algorithm alone. Note that by "gluing" all the subtrees of $\{\mathcal{T}_j\}_j$ as the corresponding leaves of $\mathcal{T}_0$, we obtain the full tree $\mathcal{T}$.

With T_j we denote the size of the tree $\mathcal{T}_j$. The set $\{\mathcal{T}_j\}_j$ we call the search tree decomposition at cut-off c, and note that it holds

$$T = T_0 + \sum_{1\leq j\leq J_c} T_j. \tag{2}$$

Next, we briefly illustrate in terms of the search tree decomposition, on an intuitive level, in which cases speed-ups can be neatly characterized and achieved and the cases where the speed-up fails.

The classical algorithm will (in the worst case) explore the entire search tree requiring T_j steps, for each sub-tree $\mathcal{T}_j$. Note that T_j characterizes the upper bound on the query complexity of the classical algorithm. Let us for the moment assume that this roughly equal to the overall run-time of the classical algorithm (we will discuss shortly when this is a justified assumption).

Further, assume that the quantum algorithm achieves a pure quadratic speed-up in run-time over the classical algorithm. Then the hybrid algorithm will take $T_H = O^*(T_0 + \sum_{1\leq j\leq J_c}\sqrt{T_j})$ time (queries), where $O^*(f(x)) \triangleq O(f(x)\cdot \mathsf{poly}(x))$ to ignore polynomial factors. To achieve a genuine speed up in the query complexity (from which we will be able to discuss total complexity), it must hold that T_H is upper bounded by $T^{1-\epsilon}$ for some ϵ (see Definition 2.3), where $T = \sum_{k=0}^{J_c} T_k$ is the total tree size.

Now this actual speed-up clearly depends on the cut-off points, and the structure of the original tree, as nothing a priori dictates the relative sizes of all the sub-trees. We first consider the case where the tree is a balanced, complete binary tree: The tree size is exactly 2^n-1 where n is the tree height. Further, assuming that the quantum algorithm can handle γn-sized instances (in terms of this natural instance size) on the κn-sized device, we get the following decomposition: $T_j = 2^{\lfloor \gamma n\rfloor}$, $j>0$, $T_0 = 2^{n-\lfloor\gamma n\rfloor}$. For simplicity, we shall ignore rounding, with which we obtain:

$$\begin{aligned} T_H &= T_0 + J_c\sqrt{T_1} && (3)\\ &= 2^{n-\gamma n} - 1 + 2^{n-\gamma n}(2^{\gamma n/2}-1) && (4)\\ &\approx 2^{(1-\gamma/2)n} && (5)\end{aligned}$$

It is clear that the above example constitutes a genuine speed up with $\epsilon = \gamma/2$, with a full quadratic speed-up when $\gamma = 1$, i.e. when we can run the entire tree on the quantum computer.

At this point it also becomes clear how the space efficiency of the quantum algorithm comes into play: If it is linear in the natural size n, then γ is a fraction of κ, and we obtain threshold-free asymptotic speed-ups. However, if the space required is super-linear in n, γ becomes a decaying function of n, in which case all speed-ups vanish. We shall briefly discuss these aspects shortly, and for more on the space complexity constraints see [1, 2]. Here we first focus on the issues stemming from the tree structure alone.

We can equally imagine a tree where $\mathcal{T}_0$ is a full tree of size $2^{n/2}$, and $J_c = 1$ with another full tree $\mathcal{T}_1$. In this case the run-time (more precisely, query complexity) of the classical algorithm is $2\times 2^{n/2} = O(2^{n/2})$, but so is the quantum query complexity: $2^{n/2} + 2^{n/4} = O(2^{n/2})$, so no speed-up is obtained. In what follows, we define J_ℓ to count all the leaves of $\mathcal{T}_0$. This is to take into account that not only do the individual sub-trees need to be large, but also need to be many in number, compared to $\mathcal{T}_0$ itself, which can be bounded by the number of leaves.

[10] It is conceivable that a sub-instance, as defined by a classical backtracking algorithm, for some reason takes more quantum space than the overall problem, and our formalism can be adapted to treat this case. However, in the cases of all quantum algorithms which we consider here, this will not be the case, hence we focus on this more intuitive scenario.

In the next section, we make the above structural-only considerations fully formal.

3.2 Criteria for speed-ups from tree decomposition

In the following, we assume to have access to a quantum computer of size $m = \kappa n$, we consider a classical backtracking algorithm $\mathcal{A}$, which generates a search tree $\mathcal{T}$. Through the study of a particular tree, one can extrapolate to a family of trees induced by the problem.

We imagine designing a hybrid algorithm based on the classical algorithm $\mathcal{A}$, and a quantum algorithm $\mathcal{A}_Q$ used when instances are sufficiently small.

We now need to characterize space and query complexities of the classical and quantum algorithm in terms of the properties of the trees.

Query complexity While the query complexity of the classical algorithm $\mathcal{A}$ is exactly the size of the search tree of the problem instance, the query complexity of the quantum algorithm may be more involved. In the case of quantum backtracking, it is a function of the tree size and height (variable number). The situation simplifies when the trees are large enough (superpolynomial in size), because the query complexity then becomes equal to the square root of the tree size up to polynomial factors.

In the case of Grover-based search, an upper bound can established by brute forcing all branching decisions along all tree paths of maximum length n. Therefore, the query complexity is exponential in the max-branching number, which in general has no simple relationship to the tree size.

Further, we highlight that the connection between the query complexities overall run-time will not aways be possible, except when:

1. the complexities of the subroutines realizing a query are assumed to be polynomial in n,
2. the query complexity is exponential, so polynomial factors can be ignored.

We embed these assumptions in our main result.

Space complexity The space complexity of $\mathcal{A}_Q$ determines the cut-off points in the search tree. The space complexity may not be a simple function of the tree height, as it depends on how the vertices are represented in memory (for instance as satisfying assignment or as the branching choices in the case of a low maximum branching number).[11]

To achieve clean polynomial speed-ups, as discussed previously, two main factors must conspire:

- the space complexity must yield a tree search decomposition where many of the subtrees which will be delegated to the quantum computer are large enough for a substantial advantage to be even in principle possible,
- and the quantum algorithm must actually realize such a polynomial advantage.

One thing to highlight in the above is the observation that the sub-trees $\mathcal{T}_j$ must not only be large on average, they need to be numerous, relative to the total size T_0 of the top of the tree. In order to take both the average size and the amount of sub-trees $\mathcal{T}_j$ into account, we do the following: Instead of considering the sub-trees $\{\mathcal{T}_j\}_{j=1}^{J_c}$, with $0 \leq J_c \leq J_\ell$, we instead consider there to be J_ℓ sub-trees, of which $J_\ell - J_c$ have size 0.

We can now identify certain sufficient conditions ensuring that an overall polynomial speed-up is achieved.

Theorem 3.1. *Suppose we are given the algorithms $\mathcal{A}$ and $\mathcal{A}_Q$, and a κ (the relative size of the quantum computer) inducing a search tree decomposition $\{\mathcal{T}_j\}_{j=0}^{J_\ell}$ as described above, for a problem family, such that:*

1. *(subtrees are big on average) The sizes of the induced sub-trees $\{\mathcal{T}_j\}_j$ are on average exponential in size, so $\sum_{j=1}^{J_\ell} T_j/J_\ell \in \Theta^*(2^{\lambda n})$ for some $\lambda > 0$. In particular, the overall tree is also exponential in size.*
2. *(quantum algorithm is faster) The query complexity of $\mathcal{A}_Q$ on exponentially large subtrees of size $\Theta^*(2^{\gamma n})$ is polynomially better than $\mathcal{A}$'s, i.e. $\Theta^*(2^{\gamma(1-\delta)n})$, for some $\delta > 0$. For convenience, we assume that*

[11] While Section 2.4 discussed the duality between formula and satisfying assignment representations, we will never employ the formula representation in the quantum algorithms due to the limited available memory.

in the quantum case the query complexity is given by some increasing concave function $\Phi(T') \leq T'$ (for a tree $\mathcal{T}'$), such that $\Phi(T')$ is essentially T' for small (sub-exponential) trees, and for bigger trees $\Phi(T')$ is essentially no larger than $(T')^{1-\delta}$.[12]

3. *(queries are efficient) The time complexities of the subroutines realizing one query in both $\mathcal{A}$ and $\mathcal{A}_Q$ are polynomial.*[13]

Then the hybrid algorithm achieves a genuine (polynomial) speed-up. If the conditions above hold for all κ, then the algorithm achieves a threshold-free genuine speed-up.

The assumption of the existence of the function $\Phi(T)$ which meaningfully bounds the quantum query complexity in terms of the tree size, is in apparent contradiction with our previous explanations that, in the case of Grover-based search, such trivial connections cannot always be made.

Here, we want to establish claims of polynomial improvements relative to the classical method. Since the query complexities of the classical method do depend on tree sizes alone, and since we assume that the quantum algorithm is polynomially faster (so the quantum query complexity is a power of the classical complexity), this implies that we can only consider cases where indeed the quantum query complexity *can* be meaningfully upper bounded by some function of the tree size. Consequently our theorem will only apply to the Grover case when the trees are sufficiently large (of size $2^{n/2}$ and higher), where such a non-trivial bound can be established.

The intuition behind the theorem statement is as follows. Considering exponentially large search trees together with the limitation on the run-times in Item 2 of subroutines to be efficient ensures that we can safely consider only query complexities to establish polynomial separations.

Secondly, Item 1 also ensures that the quantum algorithm will need to process sufficiently large problems to achieve a speed-up. This prohibits, e.g. the counterexample we gave earlier with only one tree $\mathcal{T}_1$ of size $2^{n/2}$ – the top tree has height $n/2$, meaning $J_\ell = 2^{n/2}$, and thus also the average quantity $\sum_j T_j/J_\ell$ is in this case exactly 1.

Finally, item 3 ensures that the quantum algorithm will be polynomially faster on the non-trivial subtrees. We of course implicitly assume that both the classical and the quantum algorithm solve the same problem, so that their hybridization also solves the same problem.

[12]To ensure this, we can imagine the algorithm switch from a classical to a quantum strategy only when the quantum strategy becomes faster.

[13]In fact, sub-exponential would suffice for our definitions, but reasoning is easier with polynomial restrictions

Proof. Under the conditions of the theorem it will suffice to show that the classical query complexities given by $T = T_0 + \sum_j T_j$ and the hybrid query complexity T_H, which we determine shortly, are polynomially related.

We make use of the following Lemma:

Lemma 3.2. *For any binary tree $\mathcal{T}$ of size T, the number of leaves K is bounded as follows:*

$$(T/n + 1)/2 \leq K \leq (T + 1)/2.$$

Proof. Consider T' to be the size of the binary tree $\mathcal{T}'$ obtained by replacing every sequence of one-child nodes by a single node in $\mathcal{T}$. Note that $\mathcal{T}'$ is now a full binary tree (with 0 or 2 children), yet the number of leaves of T' and T are the same. By construction we see that $T' \leq T \leq n \cdot T'$. Since full binary trees have $(T' + 1)/2$ leaves, we have that T has more or equal to $(T/n + 1)/2$ leaves, and less than $(T + 1)/2$ leaves. □

Let J_ℓ be the number of leaves of $\mathcal{T}_0$ (which includes roots of subtrees $\mathcal{T}_j$, as well as actual leaves of $\mathcal{T}$). Let $\text{Avg}_c = \sum_j T_j/J_\ell$ denote the average tree size, and by assumption $\text{Avg}_c = \Theta(2^{\lambda n})$, so that

$$T = T_0 + J_\ell \times \text{Avg}_c$$

Since J_ℓ equals the number of leaves of $\mathcal{T}_0$, by assumptions and Lemma 3.2 we have that

$$T_0 + \text{Avg}_c \frac{T_0/n + 1}{2} \leq T \leq T_0 + \text{Avg}_c \frac{T_0 + 1}{2} \quad (6)$$

which can be simplified to,

$$T_0 \text{Avg}_c/(2n) \leq T \text{ and } T \leq T_0(1 + 2\text{Avg}_c).$$

On the other hand, by similar reasoning, the hybrid query complexity T_H is upper bounded by

$$T_H \leq T_0(1 + 2\text{Avg}_q), \quad (7)$$

where Avg_q is given with $\text{Avg}_q = \sum_j \Phi(T_j)/J_\ell$ where $\Phi(T_j)$ is the quantum query complexity on the j^{th} sub-tree.

By considering the ratio of the lower bound on the classical runtime $T_0 \cdot \mathrm{Avg}_c/(2n) \leq T$, and an even weaker upper bound on the hybrid complexity

$$T_H \leq T_0(2+\epsilon)\mathrm{Avg}_q,$$

$$T/T_H > \mathrm{Avg}_c/\mathrm{Avg}_q(1/(2n(2+\epsilon))),$$

we see that polynomial improvements are guaranteed whenever Avg_c and Avg_q are polynomially related (since they are both exponentially sized by assumption, the prefactors linear in n can be neglected).

Since we have that $\mathrm{Avg}_c = \sum_j T_j/J_\ell$ and $\mathrm{Avg}_q = \sum_j \Phi(T_j)/J_\ell$, where $\Phi(x)$ is concave and increasing and $\Phi(T_j) \leq T_j$, to lower bound the speed-up characterized by $\mathrm{Avg}_c/\mathrm{Avg}_q$, we need to minimize this quantity, which is equivalent to maximizing $f(T_j) = \sum_j \Phi(T_j)$, subject to constraints $\sum_j T_j = c$, for some constant c, and $T_j \geq 0$.

By concavity of Φ, this maximum is obtained when $T_i = T_j = \mathrm{Avg}_c$ for all i, j, hence in the worst case we have $\mathrm{Avg}_q = \sum_j \Phi(\mathrm{Avg}_c)/J_\ell = \Phi(\mathrm{Avg}_c)$. By assumptions, this means $\mathrm{Avg}_q \in \Theta(2^{\lambda(1-\delta)n})$ whereas $\mathrm{Avg}_c \in \Theta(2^{\lambda n})$, which is an overall polynomial separation, as stated. This polynomial separation in the query complexity will also be observed in the final separation of the classical and hybrid runtimes because Condition 3 ensures that queries take only polynomial time.

Finally, the threshold free speedup is obtained if the above separation is obtained for any κ, which means that the tree decomposition induced by κ (and the space complexity of $\mathcal{A}_q$) yields polynomial separation between classical and hybrid runtimes. □

In the discussion above we only touch upon the sizes of search trees, without considering the possibility that the classical (online) algorithm may get lucky. Indeed, it may encounter a solution much earlier in the search tree, whereas the quantum algorithm always explores all the trees. In essence, the above considerations are for the worst cases for the classical algorithms (e.g. when no solutions exist). However, this is easily generalized.

In Section 2.4, we discussed a method to circumvent this issue [23], by enabling the quantum algorithm to only explore the effective tree, which the classical algorithm would explore as well, before termination. By utilizing these algorithms, all the results above remain valid in all cases, except the concepts of search trees must be substituted with the concepts of "effective search trees," which are the search trees the classical algorithm would actually traverse before hitting a solution.

Additionally, we observe that our framework can be generalized to allow for p-ary trees instead of binary trees.

The presented theorem has the advantage of being very general, but the major drawback is that it is non-quantitative and difficult to verify. This can be improved upon in special cases.

3.3 Space complexity, effective sizes, and tree decomposition

Three aspects complicate the task of determining the runtime of our hybrid algorithms: tree structure, time-complexity of the quantum algorithm and the space complexity of the quantum algorithm. All three issues are individually nontrivial. The tree structure can be very difficult to characterize (as is the case for the DPLL algorithm, see Section 5), and the space and time complexities can depend on different features of the sub-instance. Yet, their interplay is what determines the overall algorithm.

We address these three issues separately, focusing on settings where the situation can be made simpler. First and foremost, we will almost always assume that we deal with exponentially large trees and query complexities, and in all the cases we will consider, the runtimes involving a single query will be polynomial (Condition 3 of Theorem 3.1). For this reason, we can ignore polynomial factors, which implies that the query complexities and overall run-times are equated.

Next, since we deal with tree search algorithms, we focus on quantum algorithms $\mathcal{A}_Q$, obtained by either performing Grover-based search, or quantum backtracking over the trees. For concreteness, we will imagine the search space to be one of partial assignments to a Boolean formula (so strings in $\{0, 1, *\}$, with $*$ denoting an unspecified value), although all our considerations easily generalize. We refer to the *natural representation*, which is the vertex representation where to each vertex we associate the entire partial assignment (i.e. n symbols in $\{0, 1, *\}$, so $\log_2(3)n$ bits).

In this setting, the possible speed-ups and the space complexity can be more precisely characterized in the terms of the tree structure. In what follows, we focus on a concrete subtree representing a subproblem associated with the (restricted) formula F that $\mathcal{A}_Q$ should solve. This subtree $\mathcal{T}$ is of height n and maximum branching number $br(\mathcal{T})$, comprising partial assignments of length n.

Query complexity Recall from Section 2.4, that the query complexity of backtracking is essentially $\tilde{O}(\sqrt{T}n)$, and therefore depends strongly on the tree size and the tree height. So for large trees, we have an essentially quadratic improvement over the classical search algorithm. In the case of Grover's search algorithm, the query complexity is in $O(2^{br/2})$, provided that the maximum branching number is $br(\mathcal{T}) \leq n$, which is also a quadratic genuine speedup if the tree is full.[14]

We can already highlight the important feature that the natural instance size (number of variables) may be quite unrelated to the actual features of the sub-instance which dictate runtimes: tree size and branching number. For this reason, we introduce assumptions which allow us to relate the *variable number* with the tree size, as well as the *branching number*.

This difference between natural and effective problem sizes is even more important in the case of space complexity, where speed-ups may be impossible if the wrong measure is considered.

Space complexity We can separate two sources of memory requirements. The first is the specification of the search space, as for any methods of quantum search we require a unique representation of every vertex in the tree. Clearly, the natural representation of full partial assignments suffices, but often we can work with the specification of only the choices at every branching point.

This more efficient representation, which associates to each vertex v the unique branching choices on the path from the root to v, we refer to as the *branching representation* (see Fig. 1). The branching representation requires

[14] And, more generally a polynomial speed up when the tree size is at least $\Omega^*(2^{br/2+\delta})$, for some $\delta > 0$.

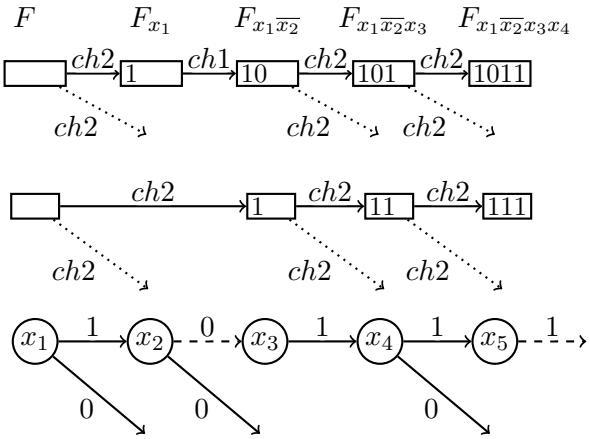

Figure 1: A path in a search tree in natural representation (above) and in branching representation (middle) for $F = (\overline{x_1} \vee \overline{x_2}) \wedge (\overline{x_3} \vee \overline{x_4} \vee x_5)$ and the corresponding path in the decision tree of a DPLL algorithm execution (below), where dashed lines are forced and solid lines guessed assignments. The node 111 in branching representation represents the partial assignment $x_1, x_2, x_3, \overline{x_4}$.

no more than $br(\mathcal{T}) \leq n$ trits or, more efficiently, $br(\mathcal{T}) + \log(n) \leq n$ bits. Technically, we need $\log(n)$ bits to fix at the depth at which our path terminates to specify a vertex (after the last branching choice there can still be a path of non-trivial length remaining to our target vertex).[15] This is still larger than the information-theoretic limit $\log_2(T)$, which is achieved by some enumeration of the vertices; however this representation is difficult to work with locally, and difficult to manipulate space-efficiently. This memory requirement is the only one which we cannot circumvent for obvious fundamental reasons.

We assume access to a function $chNo(v)$ which returns whether v has one or two children, as it was introduced in Section 2.4 along with the functions $ch1(v)$ and $ch2(v, b)$ which return the children of forced and guessed nodes respectively. We assume that for each vertex we know the level it belongs to, so there is no need to check if a vertex is a leaf. Such functions take as input a vertex specified in the natural representation, as is the case in most algorithms. The construction of functions which take the branching representation on input will require additional work and space.

[15] Our focus is not on vertices per se, but on uniquely specified paths from vertex to root, along the path of which we look for contradictions and satisfying assignments, so that the $\log(n)$ specification is not needed.

To understand the second source of memory consumption we need to consider in more detail what each search method entails.

We begin with Grover-based search. In the natural representation, if the position of the leaves is unknown, we need to perform a brute force search over all possible partial assignments, leading to the query complexity of $\Omega(2^{\log_2(3)n/2})$. This is prohibitively slow.[16]

A more efficient approach relies on the branching choice representation. In this case, the search space is 2^{br},[17] and all that is required is a subroutine which checks whether a given sequence in this representation leads to a satisfying assignment.

In other words, we require a reversible implementation of the search predicate P which is defined on branching choices.

In the natural representation, a reversible version of P for a node $\vec{x} \in \{0,1,*\}^n$ can be implemented by a circuit computing the number of satisfied clauses in $F_{|\vec{x}}$, which requires the values of the actual variables. So, in the branching representation, to evaluate P, we must in some (implicit) way first reconstruct the values of the variables that occur in the partial assignment corresponding to the vertex fixed by the branching representation. Such a modified P which evaluates the branching representation can be run reversibly, as the Grover predicate to realize the search over all possible branch representation strings. We will refer to this string as the *advice string*. Intuitively, it advises the search algorithm on its branching choices.

Now, to translate the branching representation to variable values, again intuitively, we must follow the path in the search tree, and for this, we will utilize the (reversible) implementations of the operations $ch1(v), ch2(v,b), chNo(v)$, to trace the path. This is easily done reversibly if we are allowed to store each partial assignment along the path. However, due to the size constraints imposed in the present work, realizing an efficient predicate is non-trivial task, which can nonetheless be achieved in specific cases (see Section 4).

In the case of backtracking-based algorithms, we also require sufficient space to represent each vertex uniquely. Aside from this, the space requirements stem from the implementation of the walk operator, and from the implementation of the quantum phase estimation subroutine, see Appendix B.2. The latter cost we prove can be done logarithmically in the natural representation size, and therefore can safely be ignored.[18]

We identified two subroutines as the bottleneck for a space-efficient realization of the quantum walk operator. The first is the construction of a unitary which takes a vertex specification on input (and an appropriate amount of ancillas initialized in a fiducial state), and produces (the same vertex, due to reversibility), and its children. The second subroutine is the same as in the Grover-based case: we need means of detecting whether a vertex satisfies the predicate P in whichever representation it is given.

In principle, the subroutines which detect whether a vertex is satisfying in some representation may be very different than subroutines which generate children specifications. However, in every case we discuss in this paper, the detection of satisfying vertices given in the branching representation will be implemented by essentially sequentially going through the entire path, in an appropriate representation. For this reason, the methods that we present for Grover-based search in the branching representation can readily be used for backtracking-based search. We note that the sizes of the natural, and the branching representation constitute two main parameters, *effective size measures*, associated with a problem instance, which determine the quantum space and time complexities.

In the present work, we design time-efficient schemes which achieve *linear space complexities* in the size of either the natural representation size or in terms of the branching representation sizes *for all the routines discussed above.*

[16] Furthermore in some cases it may also lead to invalid results, if there is no explicit mechanism to recognize legal vertices in the tree, and if the connectivity encodes properties of the problem.

[17] Note, this does not enumerate all the vertices or leaves, but possible paths in trees with no more than br branchings. This suffices to uniquely specify a leaf. There may be multiple specifications for a single leaf, if it occurs on a path with fewer branchings (in which case, the remaining choices are then simply ignored).

[18] Note, a linear dependence would not immediately prohibit the application of a hybrid method, but it would cause a multiplicative decrease in the effective size of the instance we can handle, i.e. our usable work space would effectively become a fraction of what it could be.

Also they can be applied both in the Grover-based and backtracking cases. Additional sub-linear space contributions are effectively negligible: since we assume quantum computers that are proportional the instance size, so κn, so we can simply sacrifice any arbitrary sized fraction ϵn of κn (so decreasing κ by an arbitrarily small ϵ) for all sub-linear space requirements.

With a full understanding of what parameters of the sub-instances influence space and query complexities of the quantum algorithms we consider, we can now focus on the overall tree decomposition, and identify settings where it all can be made to provide simple criteria for quantifiable speed-ups.

3.4 Speed-up criteria: special cases

In this section, we introduce several assumptions to simplify the expressions of the runtimes obtained through our hybrid approach. One of the main assumptions is that the effective size of the quantum computer can be expressed as $\kappa' n'$, given a $\kappa n'$ sized device; in other words, that the space complexity of our algorithms is linear in n', where n' quantifies the original problem size measure. In the beginning of this section n' will refer to the natural instance size (and thus the tree height), but later we will show how it can also quantify branching numbers.

Given a $\kappa' n'$ effective size quantum computer, the search tree decomposition will then produce a cut-off points at each vertex where the vertex effective size n (and the branching number br) is below $\kappa' n'$.

To achieve polynomial speed-ups resulting subtrees must be exponential (on average, relative to the number of leaves of the top tree $\mathcal{T}_0$), i.e. in $\Theta^*(2^{\lambda n})$, where speed-ups are achieved using backtracking for any λ, and using Grover if $\lambda > 1/2$, or if $\lambda > br/2$ when the branching representation is used.

To make this more concrete we can consider a setting where exponential subtree sizes are guaranteed and easily computable, i.e. when the subtrees are uniform at a scale determined by $\kappa' n$.

For the weakest case, where the space complexity of the quantum algorithm we consider is governed by n (in the natural representation), we have the following setting.

We will say that a fully balanced[19] tree $\mathcal{T}$ is *uniformly dense with density larger than λ at scale ηn*, if all the sub-trees $\mathcal{T}'$ of height higher than ηn (i.e. all trees with a root at a vertex v of $\mathcal{T}'$ at a level higher than $n - \eta n$, which contain all the descendants of v at distance no more than ηn from v) satisfy $T' \in \Theta(2^{\lambda \eta n})$.

In this case, if η is matching the effective size of the quantum computer ($\eta = \kappa'$), we get a polynomial speed-up for all densities λ for backtracking, and whenever $\lambda > 1/2$ for Grover-based search, assuming access to a quantum computer with effective size $\kappa' n$ (with κ' proportional to κ). The assumption that the trees are exponentially sized makes the proof of this statement trivial.

The definition of such strictly uniform trees can be further relaxed, by allowing that just a $1/\mathsf{poly}(n)$ fraction of the sub-trees beginning at level $n - \eta n$ is exponentially sized with the exponent λ, and by somewhat freeing the size of the top tree $\mathcal{T}_0$ (we essentially only need that this tree is not too large), and still obtain a polynomial improvement. Again if $\eta = \kappa'$, we get polynomial speed-ups when κ' is proportional to κ.

Since the effective size is $\kappa' n$, for uniform trees we obtain a search tree decomposition, where we cut at level (instance size given by height) $n - \kappa' n$. The obtained trees are of size height $\kappa' n$, and by assumption, $1/\mathsf{poly}(n)$ of the trees are exponentially sized (i.e. are in $\Theta^*(2^{\kappa' \lambda n})$). But then the worst case average size is given with $2^{\lambda \kappa' n}/\mathsf{poly}(n)$ which is still exponential.

With this we satisfy all the assumptions of the Theorem 3.1 and conclude polynomial improvements. But in this case we can be more precise about the achieved improvement. From the proof of Theorem 3.1, Eq. (7) the hybrid query complexity can be upper bounded with $T_0(1+2\mathrm{Avg}_q)$ where Avg_q is given with $\mathrm{Avg}_q = \sum_j \Phi(T_j)/J_\ell$.

Here $\Phi(T_j)$ is the quantum query complexity on the j^{th} sub-tree. Note that we assume that we can meaningfully bound the quantum query complexity as some function of the tree size. If we are using backtracking since the trees are exponentially large we have that $\Phi(T_j) = \Theta^*((T_j)^{1/2})$. In the case of Grover's search, we will have meaningful statements of this type when the tree sizes are exponential in depth (so $T_j = \Theta(2^{\lambda n'})$), with exponent $\lambda > 1/2$.

[19] All paths from the root to a leaf are of length n.

By the same proof, Avg_q is maximized when $T_i = \sum_j T_j / J_\ell$, for all i, in which case

$$\text{Avg}_q = \Theta(2^{\lambda\kappa' n/2}/\text{poly}(n)),$$

which is dominated by $\Theta(2^{\lambda\kappa' n/2})$.

$$T_H = O^*(T_0(2^{\lambda\kappa'/2n}) \quad (8)$$

$$= O^*(2^{(\kappa'/2+(1-\kappa'))\lambda n}) = O^*(2^{(1-\kappa'/2)\lambda n}) \quad (9)$$

which is a polynomial improvement over the classical strategy which obtains

$$T_H = \Omega^*(2^{\lambda n}) \quad (10)$$

(recall, the asterisk denotes we omit polynomial terms).

As noted earlier, in the above analysis, if we use Grover's search, then

$$\Phi(T_j) = \min(2^{\lambda\kappa' n}, 2^{\kappa' n/2}),$$

in which case we only obtain a speed-up if $\lambda > 1/2$.

However, in Subsection 4.2 we describe a setting where, although the tree is quite uniform, it is not sufficiently uniform relative to the natural measure – the tree height. However, it is uniform relative to the branching number measure. We can easily adapt the uniformity definition to this case.

We will say that a tree $\mathcal{T}$ is *uniformly dense at branching level* ηn, if all the sub-trees $\mathcal{T}'$ for which we have $br(\mathcal{T}') \geq \eta n$ (i.e. all trees with a root at a vertex v of $\mathcal{T}$ which have at least ηn branches on any path) satisfy $T' \in \Theta(2^{\lambda\eta n})$.

Now, if we have access to algorithms whose space complexity is linear in the branching size measure, given a $\kappa' n$ effective quantum computer size – now, relative to the branching size, meaning we can handle instances with $\kappa' n$ branchings –, if the trees are uniformly dense at branching level ηn, with $\eta n \leq \kappa' n$, we can run the quantum subroutines on exponentially sized subtrees and a similar analysis holds. But this time, Grover's approach yields speed-ups whenever $br/2 < \lambda$.

Although these results seem very similar, in general the latter setting allows us to start search much earlier as $br \leq n$, and indeed, it can be much smaller. Furthermore, in some cases, the distribution of branch cuts is not (guaranteed to be) uniform with respect to the natural size n, which prevents naive algorithms to be successful.

In the next sections, we provide examples of both cases:

- an example where the effective size n plays a role, in the context of k-SAT formulas for large k and uniform trees, with speed-ups obtained via Grover-based search ($\lambda > 1/2$) and backtracking (Section 4.1);
- an example where the number of maximal branches br plays a role, with speed-ups that depend only on the branching number measure (Section 4.2).

We provide in Appendix D an example where backtracking provably provides better hybrid performances than Grover for the Hamiltonian cycle problem, building on prior work [2].

4 Hybrid speed-ups for tree search in satisfiability problems

In this section, we provide examples of when various types of speed-ups are attainable using the hybrid tree-search-based framework we introduced in previous sections.

4.1 Algorithm-independent improvements under the Strong Exponential Time Hypothesis

The simplest and most ideal setting for hybrid divide and conquer approaches is when the search trees of any SAT algorithm are essentially everywhere maximally dense. The Strong Exponential Time Hypothesis (SETH) implies this. Under this hypothesis, we can provide the simplest hybrid algorithm which still beats the best possible classical algorithm for k-SAT (where k is large).

Let us write γ_k for the smallest $\gamma_k \in [0, 1]$ such that there exists a k-SAT randomized algorithm of complexity $O(2^{\gamma_k n})$. SETH stipulates that the sequence of all the γ_k's is increasing, and $\lim_{k\to\infty} \gamma_k = 1$. Then γ_k can be made arbitrarily close to 1. In other words, this also means the best possible classical algorithm is close to brute-force search, which is itself a divide-and-conquer algorithm, for large k.

We can apply the hybrid approach to brute-force search as classical algorithm, and Grover's search as quantum subroutine, with access to a κn-qubits quantum computer. As we show in the Appendix B.1, we can implement Grover-based, brute-force search for SAT solving over n-variable formulas using $n + O(1)$ space, meaning that, asymptotically the effective size of the

problem we can handle is $\kappa' n$ with $\kappa = \kappa'$. The result is a hybrid algorithm of time complexity $O^*(2^{(1-\kappa/2)n})$ for k-SAT for every $k \in \mathbb{N}$.

But then by SETH, for every $\kappa > 0$ there exists a k s.t. $\gamma_k > 1 - \kappa/2$, which implies a polynomial speed-up over any classical algorithm. In summary we have the following result.

Theorem 4.1. *Under the Strong Exponential Time Hypothesis, for every classical algorithm for k-SAT and for every $\kappa > 0$ (such that we are given access to a κn-qubits quantum computer), there exists a k such that we obtain a speed-up for the hybrid divide-and-conquer algorithm based on classical brute-force search and Grover's algorithm. In other words, the hybrid divide-and-conquer approach can offer an algorithm-independent speed-up under SETH.*

To connect to the previous discussions on uniform trees, SETH guarantees that the trees of any tree-search based algorithm for k-SAT will, become arbitrarily dense (close to $\lambda = 1$), at every constant scale κn, for large enough k.

4.2 Threshold-free speed-ups in PPSZ tree search for special formulas

In this section we describe a setting in which, under some assumptions on the variable order, hybrid, threshold free speed-ups are possible for PPSZ tree search, which, as we mentioned, is at the core of the best-known classical exact SAT solvers.

4.2.1 Characterizing the search trees

Recall that, towards realizing hybrid tree search, we already introduced dncPPSZ, a tree search version of the original Monte Carlo algorithm (see Algorithm 2 in Section 2). It finds a satisfying assignment in a constant number of repetitions if one exists (see Proposition 2.1). As input, the dncPPSZ subroutine takes an k-CNF formula F and an order π, then sequentially either resolves the next variable by s-implication, yielding a forced tree node, or it branches on that given variable, yielding a guessed tree node. As detailed in the Appendix A, there exists a constant $\varepsilon > 0$ such that, if satisfying assignments exist, then $E_{\pi \in S_n}[\text{dncPPSZ}(F, \pi) = 1]$ is non-nil and constant (and thus a finite amount of repetitions of dncPPSZ suffices to find the satisfying assignment). In other words, with constant probability dncPPSZ will find a path from the root to a satisfying assignment which contains no more than $b = (\gamma_k + \varepsilon)n$ branches (guessed variables). By pruning tree paths beyond b branches, we obtain a runtime of $O^*(2^{(\gamma_k+\varepsilon)n})$ and a search tree with maximal branching number b (see Def. 2.2).

Since, our objective is to provide a hybrid algorithm which achieves a better provable upper bound, threshold-free, the first obstacle is obtaining a sufficiently dense tree shape (see Section 3.2). However, since no such algorithm is known, we indeed may assume the trees are of size $\Theta^*(2^{(\gamma_k+\varepsilon)n})$ (at least in the worst cases). This all suggests that (non-hybrid) speedups are obtainable with quantum backtracking, and with a Grover-based method as well; recall that Grover achieves the same upper-bound performance as backtracking, if the bound on the tree size is given by the exponential of maximal branching number, as discussed in Section 2.4.1. Any results with Grover characterize what we can expect from a quantum tree search algorithm.

Next, in hybrid setting we need to keep track of space complexity as well, and the depth κ' at which to invoke the quantum subroutine. Since PPSZ is a simple tree search algorithm over partial assignments, the natural instance size is the number of variables which have not yet been set. It is relatively easy to construct quantum Grover-based and backtracking algorithms which achieve a linear space complexity in this quantity. In this case, the cut-off point in the search tree decomposition happens at vertices corresponding to some number of variables not yet resolved. However this will not suffice for threshold-free improvements.

The problem is the following: while the number of branches is $(\gamma_k + \varepsilon)n$, along a path from the root to an assignment, there is no guarantee on *where* along the path they occur. Indeed, since the formula simplifies as variables are set, it is more likely branches occur earlier on, and it is possible all branches are used up first, leaving a formula with $n - (\gamma_k + \varepsilon)n$ variables, which is trivial from this point. So to achieve speed-ups in this scenario, our quantum device must be able to handle instances of size at least $\kappa' n > (1 - \gamma_k - \varepsilon)n$, hence the approach is not threshold-free.

Our solution is to construct algorithms that

work in the branching representation, discussed in Section 2.4.1, and achieve linear space-efficiency in the remaining number of branches. Since branches directly dictate the (exponential) tree size, starting the algorithm at a point with some-fraction-of-n branches remaining will guarantee all the properties we highlighted in Theorem 3.1, and more specifically, render the PPSZ case a uniform tree case relative to the branching cuts, as described in Section 3.4. However, coming up with reversible implementations of PPSZ traversal, whose space efficiency depends on the branching numbers (and time-efficiency is sub-exponential) is more complicated.

At this point let us be more precise. We are looking for reversible algorithms (circuit) which compute the children for any vertex of the PPSZ search tree in the branching representation, and which can decide whether a vertex specified in the branching representation is satisfying or a contradiction. That is, implementations of the functions $ch1, ch2, chNo, P$, as this suffices to implement the walk operator (see Section 2.4).

What complicates the realization of such subroutines is, as we discussed previously in Section 3.3, branching choices, i.e. the values of guessed variables alone, do not map trivially to individual variable values. For example, in the branching representation, a node's label 111, i.e. the first three guessed variables have all been set to 1, should first be converted into a partial assignment to the variables, before we can compute its children.[20] To know which variable is the first guessed variable, we must first compute the s-implications. An obvious solution is to compute all implications and store the corresponding partial assignment, but this violates the objective of using less memory than what the natural representation allows. Since, as mentioned, Grover-based search will already achieve speed-ups over a classical strategy, we tackle the above problem in this context.

First, we simplify the problem for the PPSZ setting by using eager evaluation of forced variables. This compresses all the line paths in the tree, resulting in a tree with only binary nodes. Algorithm 3 illustrates this using the tree node functions. The node v consists of an advice label, e.g., 111 (including a counter indicating its length). Implementing $ch2$ is trivial, e.g., $ch2(111, b) = 111b$. However, this complicates the search function P since it now has to eagerly evaluate forced variables, but it anyway needs to 'decode' the advice (which is a similar process).

Algorithm 3 dncPPSZ(v) using our functions, where $|v|$ is the length of node v's advice label.

if $P(v) = 1$ **then return** 1
if $P(v) = 0$ **or** $|v| > (\gamma_k + \varepsilon)n$ **then return** 0
return dncPPSZ($ch2(v, 0)$)$\vee$dncPPSZ($ch2(v, 1)$)

Next, we specify the problem that P has to solve. We call it *s-implication with advice* (SIA), which is intuitively and implicitly defined as follows (Appendix E provides a precise definition). Given a fixed formula F and variable order π, an algorithm solving SIA takes on input an advice string of size S_{adv}, specifying which branching choices will be made at the guessed variables. The algorithm should compute the path realized in an s-implication-resolution based process in the tree specified by F by going over the variables in the specified order, setting them to either the forced value if s-implied, or to the next unused branching choice specified in the advice register if not. It should then output: $|00\rangle$, if more branches are encountered than the advice length; $|1b\rangle$, if the path makes the formula (un)satisfied at some point, such that $F = b$.

The desired space-efficient implementation of SIA will utilize a S_{adv}-sized advice register for the choices, and ideally no-more than $o(n)$ ancillas (although, low-prefactor linear scaling is also acceptable as explained in Section 3.3). While these criteria are easily met in an irreversible computation, critically, it must be realized reversibly under these conditions, as in this case we achieve Grover-based search, by searching over the advice register.

Obtaining $o(n)$ space is not trivial for the problems we consider. In particular, it can be proven that the routines required to implement PPSZ (e.g. repeated s-implication, unit resolution) are P-complete under log-space reductions (as proven for unit resolution in [26]). It is still an open problem whether $o(n)$ space, poly-time

[20] For instance, in the formula $(\overline{x_1}\vee x_2)\wedge(\overline{x_2}\vee\overline{x_3}\vee\overline{x_4}\vee x_5)$ (and order $x_1 < \cdots < x_5$), the node with branching representation 111 represents the assignment $x_1, \overline{x_2}, x_3, x_4, \overline{x_5}$, because x_2 and x_5 will be $(s > 1)$-implied.

algorithms exists for P-complete problems [27, Section 5.4], but lower bounds on pebbling approaches [28] give a pessimistic outlook. Therefore, an approach that focuses on identifying classes of formula which admit genuine hybrid speedups seems justified.

Accordingly, we focus on formulas where, for a given variable order, the formula has *bounded index width* (biw) as defined in Definition 4.2. In Appendix E, we provide an $o(n)$-space reversible algorithm for SIA, which works for this class of formulas. Specifically, we provide an algorithm which has polynomial runtime, and space complexity $S_{adv}+S_w$, where S_{adv} is the advice-string register, and

$$S_w = O(w \cdot \log(n/w)) \tag{11}$$

for any k-CNF formula of biw w. We do so by providing a frugal, reversible implementation of SIA, and then by splitting this (deterministic) computation into blocks that require only a partial assignment of w variables, applying Bennett's reversible pebbling strategy on those blocks [29].

Definition 4.2. *For a given Boolean formula F and a variable ordering $\vec{x}$, we defined the index-width* $\mathrm{iw}(F)$ *of a formula to be the largest difference between two indices of variables in a clause of the formula F, i.e.*

$$\mathrm{iw}(F) = \max_{C\in F} \max_{x_i,x_j\in C} |i-j|.$$

A formula F has bounded index width (biw) w if $\mathrm{iw}(F) \leq w$.

Before giving the complexity theoretic analysis of the hybrid method obtained by using the SIA algorithm above, we highlight a few facts.

First, the property of bounded index width is an order-dependent property. Randomizing the variable order breaks this, and consequently our result does not directly imply speed-ups for PPSZ-proper, just for the tree-search part.

Luckily, an inspection of the correctness proof of PPSZ in [20] reveals that not all variable orders need to be considered in the case of biw formulae. This is because the biw property ensures that the critical clause tree, central to the proof, only reasons over variables up to distance $w \cdot \log(s)$ from the variable for which s-implication is computed. The critical clause tree can thus not distinguish between permutations with more than $w \cdot \log(s)$ displacement. Permutations with no more than k displacement can be efficiently obtained by k-sorting[21] a random permutation [30]. A k-sorted permutation of the variable order partially preserves the biw: if a formula F has biw w for the initial variable order, then a k-sorted permutation of the variable order yields a biw $w' \leq w+2k$.

Second, bounded index width formulas have specialized SAT-solving algorithms with best run-time $O(2^w\mathsf{poly}(n))$, to our knowledge [31]. We will discuss the consequences shortly but for the moment we just focus on beating known PPSZ tree search bounds in these special cases. Next we continue with complexity theoretic analysis of the hybrid algorithm with SIA.

Consider first the setting where the index width is sub-linear, $w \in o(n)$. In this case, the space complexity is sub-linear (barring the advice), which means that given a κn-sized quantum computer, we can turn to the quantum strategy the moment $\kappa' n$ guesses remain with $\kappa' > \kappa+\epsilon$, for every $\epsilon > 0$ (we reserve ϵn memory to hold the $o(n)$ ancillas).

From this point on, we instantiate the discussion regarding uniform trees with respect to branching cuts discussed in section 3.4.

We can assume that the sub-trees are exponentially sized, of size $\Theta^*(2^{\kappa' n})$ (as this quantity is used as the upper bound on the time complexity of the classical algorithm, which we are trying to outperform), and we obtain a full quadratic speed-up run-time of $O^*(2^{\kappa' n/2})$ on the same subtrees.

The top part of the tree has size $O^*(2^{(\gamma_k-\kappa')n})$ (as tree sizes are dictated by branching choices), and by our previous analyses, this yields a hybrid run-time of $O^*(2^{(\gamma_k-\kappa'/2)n})$ (when $\kappa' < \gamma_k$, otherwise we obtain a full quadratic speed-up). All in all, in terms of κ, we obtain $O^*(2^{(\gamma_k-(\alpha\kappa-\epsilon)/2)})$, were $\alpha = \kappa'/\kappa$ is the coefficient from the space efficiency of SIA, and ϵ is an arbitrary small constant used to handle $o(n)$-sized ancillary registers.

We note that interesting examples of bounded index-width formulas (with connections to statistical physics) arise when one consider restrictions of the 3-SAT problem. One example of such problem is Lattice SAT (3-SAT on a lattice, with $\sqrt{n} \in o(n)$ bounded index width), which is formally defined and proved to be NP-complete in

[21] A sequence $[a_1,\ldots,a_n]$ is k-sorted iff $\forall i,j$ with $1 \leq i \leq j \leq n$, $i \leq j-k$, it holds that $a_i \leq a_j$.

Appendix G.

At this point, we return to the fact that for bounded index width formulas specialized SAT-solving algorithms exist with best run-time of $O(2^w \mathsf{poly}(n))$ [31]. In the case that $w \in o(n)$, we can decide the very satisfiability of the given formula in subexponential time.

In other words, in the PPSZ process, it is much more efficient (in terms of upper bounds) to actually solve SAT the moment we encounter a formula with a sub-linear index width, than to continue with PPSZ search. Switching from the tree-search process to a specialized solver is of course no longer PPSZ, but a new algorithm whose properties are uncharacterized (but not worse than PPSZ). Consequently, technically, our previous results still entail an improvement over the basic PPSZ tree search. However, it is of course interesting to see if settings can be identified where switching to a bounded-index-width specialized algorithm actually constitutes a bad choice, and our hybrid strategy is best in general. We provide this in the next section.

4.2.2 When hybrid quantum PPSZ search improves over known classical algorithms

To defeat bounded-index-width specialized algorithms, we consider index widths which are linear in n. As shown in Appendix E, it is possible to implement s-implication with advice (SIA) reversibly using total space $S_{adv} + S_w$, where $S_w = O(w \cdot \log(n/w))$ (the ancillary space needed for reversible SIA), and where S_{adv} is the size of advice itself. For all w, the runtime of this subroutine is polynomial in n, and the runtime of the overall quantum algorithm, which solves the subtree by exploring the advice string space, is $O(2^{S_{adv}/2} \times \mathsf{poly}(n))$.

In contrast, a classical algorithm can decide SAT on formulas of index width w in time $O(2^w \mathsf{poly}(n))$ [31]. It is unlikely that one can do much better than this: in the case $w = n$, achieving a bound better than $O^*(2^{cn})$, for some constant $c > 0$, would violate the exponential time hypothesis (ETH).

Here we assume this also holds for index widths which are fractions of n.[22] It will be convenient to fix $w = \zeta n$ (note ζ is a constant). In this case, we can identify a regime in which the quantum search over the advice string is still faster than solving SAT on the formula. This is the case whenever $S_{adv}/2$ is less than w, as these are the exponents of the exponential part of the run time of the respective algorithms. However, we must still ensure that the quantum algorithm can be run at all. So, a part of the overall memory available must be split between the advice string, and the memory required to solve SIA with advice.

We set $S_{adv} = \beta\kappa n$ for $\beta \in (0, 1)$, and the remainder of the memory of $(1 - \beta)\kappa n$ qubits is spent as ancillary memory S_w for the SIA subroutine. Note that when setting $\zeta = w/n$ as a constant, we have that $S_w \notin o(n)$, as w is then linear in n. However, in what follows we show that we can always find a pair (β, ζ) such that at least $S_w < n$.

In summary we have that

$$\log(1/\zeta)\zeta b \leq (1 - \beta)\kappa \tag{12}$$

since to process $w = \zeta n$-index width formula we need $\log(1/\zeta)\zeta bn$ bits (for some constant b; Eq. (11)), and since we allocated $(1 - \beta)\kappa n$ for this purpose. Moreover, the exponents of the run-times of the quantum and classical algorithms are $\beta\kappa n/2$ (Grover speed-up) and ζcn, respectively, so it must hold that

$$\beta\kappa < 2c\zeta. \tag{13}$$

For our purposes, it suffices to show that for every κ,b,c, there exist a (β, ζ) pair for which both conditions hold.

First, we fix a β, to some (small) value β'. Next, find ζ satisfying Eq. (12); note such a ζ exists as $\log(1/\zeta)\zeta$ decreases in ζ, when ζ is small enough, converging to zero. Note, if Eq. (12) holds for $\beta = \beta'$ it also holds for any smaller β. Then choose $\beta \leq \beta'$ such that Eq. (13) is satisfied, which can be done by choosing β small enough. Since these (β, ζ) exist for any b, the space overhead of S_w can be made an arbitrarily small (constant) fraction of n.

This guarantees the existence of regions in the space set by the advice size (controlled by β) and index width (controlled by ζ) with polynomial speed-ups (for a given κ). However, there are no guarantees that the PPSZ process is guaranteed to generate formulas which will fall in this region, as discussed shortly.

[22] Furthermore, under the strong ETH, c approaches 1 for large clause sizes (k in k-SAT).

4.2.3 Grover vs quantum backtracking

We note that, instead of running Grover's search over an advice string, the same algorithms can be utilized to perform quantum backtracking over the same trees, as we briefly announced previously. The backtracking tree corresponds to restricted formulas as usual, where nodes have two children (applying s-implication eagerly to collapse forced nodes with one child).

Determining children corresponds to the step SIAB (see Appendix F), and the evaluation of the leaves (final satisfiability) will actually involve running the entire scheme developed for the Grover-based approach. The advantage here is that we will only explore the actual tree, but this does not provide a better theoretical speed up as the classical bounds assume a full tree.

However, the downside is that the walk operator utilizes two copies of the search space, and search-space-sized ancillary register, so that the available space is reduced by a factor of 4 (see the construction of the walk operator in Appendix C.3). This nonetheless allows for regions of threshold-free polynomial speed-ups (κ would be replaced with $\kappa/4$ in the above analysis). This would lead to smaller improvements in the upper bounds (i.e. ignoring the speed-ups from exploring smaller trees), but may overall be more efficient in practice. We note that this pre-factor of 4 can probably be further improved.

4.2.4 Assumption on the variable order

We now discuss two clashing requirements from previous subsections. First, to allow for a space-frugal reversible implementation of s-implication, in Subsection 4.2.1 we consider formulae with bounded index width w, and permutations of these formulae with biw $w' = w + 2w \cdot \log(s)$. Second, to obtain a speedup over classical algorithms for biw formulae, in Subsection 4.2.2 we require that w grows linearly in n. This gives rise to an issue with regard to the space requirement: with w linear in n and s growing very slowly in n, we have that the ancillary space requirement (Eq. 11) for the permuted variable orders grows as $O(w \cdot \log(s))$, which grows faster than $O(n)$.

The permuted variable orders are used in the PPSZ proof to guarantee an upper bound of $\gamma_k n$ on the expected number of guessed variables during a PPSZ run [20]. Limiting the k-sorted permutations to some smaller k, e.g. $k = w$ would solve the space requirement, but fails to retain the guarantee on the expected number of guesses.

We therefore add the assumption that, for a formula F with biw w, at least $1/\mathsf{poly}(n)$ of its w-sorted permutations are "good" variable orders. With a good variable order we mean a variable order for which the expected number of guessed variables during the PPSZ run is at most $\gamma_k n$.

4.2.5 Putting it together

Under the assumption made on the variable order in the previous subsection, we can put together the hybrid PPSZ algorithm as follows below. The result is summarized in Theorem 4.3.

In a hybrid run of PPSZ-proper, which calls PPSZ tree search on order-randomized instances, the algorithm would keep track of the current restricted formula. Then, it would switch to the quantum subroutine whenever the constraints for speed-ups, depending on the advice size (which is equal to the number of guesses remaining before $\gamma_k n$ guesses have been used up), and index width can be satisfied. Unfortunately, it is not the case that all the subtrees (for all initial formulas) will necessarily end up in the regions satisfying the criteria, while the subtrees are still exponentially sized.[23] However, in those cases, we are also not dealing with a hard instance, in the sense of the exponential time hypothesis.

When the trees are large enough, we will obtain a polynomial speed-up on that subtree, but this must occur for a constant fraction of the trees, to achieve an overall polynomial speed-up. This is true even if the initial formula is of bounded-index width, which is convenient as index width can only decrease, as index width, and number of guesses remaining can decrease at different rates.

Theorem 4.3. *For a class of Boolean formulas F, the hybrid PPSZ algorithm achieves a polynomial speed-up if the following conditions are met:*

1. *F is a hard instance (in the sense of the exponential time hypothosis).*
2. *F has bounded index width w, and w grows linearly in n.*

[23] The formulas will become eventually small enough but if number of guesses decreases earlier than the index width, we may end up with trivial trees; in this case, interestingly, PPSZ will be faster than the upper bounds on index-width-specialized SAT solver.

3. *Out of all w-sorted permutations of the variables of F, at least $1/\mathsf{poly}(n)$ have the property that the expected number of guessed variables during the PPSZ run (given that permutation) is at most $\gamma_k n$.*

Similar results may be obtainable for planar formulas. This is interesting as planarity does not depend on the variable order, so will not be obstructed by the randomized order introduced by PPSZ. However for planar formulas, we have subexponential algorithms with runtime $O(2^{\sqrt{n}})$, so PPSZ is not the best choice to begin with.

Finally, note that, outside of the context of our hybrid divide-and-conquer approach, our results imply that there is a quantum implementation of PPSZ via quantum backtracking which achieves a quadratic speed-up over classical implementations of PPSZ.

Theorem 4.4. *The existence of a quantum implementation of SIA (Appendix E) together with Algorithm 3 implies a quantum backtracking implementation of PPSZ which achieves a quadratic speed-up over classical PPSZ.*

5 Discussion

The previous results constitute settings where we could obtain speed-ups for well characterized cases. In this discussion section, we consider the applicability of the hybrid method to the DPLL algorithm, and briefly discuss the consequences of polynomial time cut-offs, and alternative scalings, and finally, the limitations of hybrid methods.

5.1 Potential for DPLL speed-ups

In [3], Montanaro has demonstrated how quantum backtracking can be used to speed-up basic DPLL algorithms, which utilize just unit rule and pure literal rule resolution methods. This suffices for polynomial speed-ups whenever the search trees are exponentially large, and the satisfying solution does not exist, or appears late in the search order in the classical algorithm.[24]

However, to achieve improvements in hybrid settings, with a κn-sized quantum computer, we have additional criteria on the structure of the tree and the algorithm as discussed in detail previously. This makes matters more complicated for DPLL. In general, there is less theory for DPLL we could apply to these questions in comparison to the PPSZ method. Nonetheless, we can at least resolve some of the technical concerns regarding the subroutines.

In Appendix C, we provide space-efficient implementations of quantum backtracking for pure literal rule and the unit rule, which suffice to obtain essentially linear space complexity with respect to the natural instance size (number of free variables). In this case, any set of formulas which generate dense restricted formulas at depth $\kappa' n$ (where $\kappa' n$ is the effective quantum computer size) can be sped-up in a hybrid scheme. At present, we can only state that this is guaranteed (under SETH) k-CNF formulas (for high k), as explained in Section 4.1. However, as characterizations of the lower bounds of DPLL trees improve, it is possible we obtain provable speed-ups for interesting κ ratios, if not threshold free.

Using the same methods we provided for s-implication with advice for bounded-index width formulas, we can also provide equally space efficient algorithms for a space-efficient pure literal rule with advice.

As stated earlier, we can implement all these algorithms in the quantum backtracking setting, achieving speed-ups in terms of the tree size, whereas branching number just determines the space complexity. In this case, we could employ the quantum algorithm earlier, depending not on the natural instance size, but rather, on the number of branches like in PPSZ. However, the problem is that in the case of DPLL we have no meaningful upper bounds on the needed advice size. This can cause false negatives: if the quantum algorithm is utilized too early, we will run out of advice even if there exists a path to a satisfying assignment.

Since running the search still constitutes an exponential effort (in the advice string size), we cannot simply run the algorithm at each vertex of the tree. Nonetheless, this approach may offer a path to viable heuristic for the classes of formulas where we have a reasonable upper bound on the number of branches, or additional information on the tree structure – since DPLL is predominantly a heuristic method, such a result is fitting.

We end our discussion on the hybrid method

[24] More precisely, we require that the effective explored trees, as discussed in Section 3 are exponentially sized.

for DPLL by considering an additional constraint: real world considerations, specifically that all the run-times are polynomially bounded.

5.1.1 DPLL in the poly-domain

In practice, DPLL is often used as a heuristic, on formulas for which it can find a satisfying assignment with high probability in polynomial time.

In this subsection, we consider the possible consequences of obtaining small polynomial improvements over classical DPLL with polynomial cut-offs using a hybrid, or fully quantum method. The problem is: in the poly-world, poly-overheads, which we ignored in all previous considerations, matter.

For some $c > 0$, we define a polynomial cut-off for DPLL to be a polynomial size limit n^c for the subtree explored by DPLL for an arbitrary formula. On such a subtree, classical DPLL will take $\mathcal{C} = \text{poly}_1(n) \cdot n^c$ time to terminate, while our hybrid DPLL will take $\mathcal{Q} = \text{poly}_2(n) \cdot n^{\alpha c}$, where $\alpha \in [\frac{1}{2}, 1)$ depends on the shape of the subtree defined by the first n^c vertices explored by DPLL, and the size of the quantum computer that we are given access to, and $\text{poly}_1(n)$ and $\text{poly}_2(n)$ are the run-times of individual subroutines in involved in one query of the classical and quantum algorithm, respectively.

Note that in this setting, the size T' of the subtree explored by the algorithm is much smaller than the size of the whole search tree T, and therefore, we need to exploit a variant of quantum backtracking, whose inner working is explained in Appendix C.6.

The classical runtime is greater than the hybrid divide-and-conquer runtime whenever $\mathcal{C} > \mathcal{Q}$, i.e.

$$\frac{\text{poly}_2(n)}{\text{poly}_1(n)} < n^{(1-\alpha)c},$$

so that our hybrid divide-and-conquer approach improves on the classical runtime whenever $c > \frac{\beta}{1-\alpha}$, where β is such that $\frac{\text{poly}_2(n)}{\text{poly}_1(n)} = n^\beta$.

It would be interesting to estimate how the ratio $\alpha \in [\frac{1}{2}, 1)$ evolves depending on κ, for various ordering of the vertices of the search tree (depth-first search and breadth-first search in particular).

A careful reader may notice that numerous problematic assumptions have to be taken into account to achieve, arguably, very small improvements. We point out that this is all a consequence of our setting chosen to enable us to provide *clean statements about asymptotic speed-ups*. In particular, it is for these reasons that we assume that the quantum computer scales with some quantity which can be used to characterize run times and space efficiencies. This is the "natural instance size".

However, one can easily switch to a less demanding model. Let $c(n)$ be the space complexity of the fastest known quantum algorithm for the problem class under considerations. If we assume that the quantum computer is of the size $\kappa \cdot c(n)$, which is still smaller than what we need to run the basic algorithm hence interesting, stronger results may be possible, with the expense being a less clean, and more conditional, analysis.

Note that, in the real world, a quantum computer will offer an advantage in real-compute time, whenever it is used in a hybrid setting and the quantum device achieves a real-compute-time speed-up (not as a scaling statement, but in the units of seconds), on the particular sub-instance. Real-world analyses of speed-ups, for fixed instance sizes must thus take into account real-world parameters.

Although in the present work, we are interested in the more theoretical questions, we believe that further improvements are still possible in this much more stringent, and more general model. This leads us to the next section, which discusses the general limitations on the speed-ups obtainable by the the hybrid method.

5.2 Limitations of the hybrid approach and the framework of networked smaller quantum devices

In the approaches we have explored, the best improvement one can obtain given access to a fractionally smaller quantum device is polynomial. This is not just a consequence of the fact that we use quantum backtracking or Grover's search as the backbone of the speed-up (which themselves only allow a quadratic improvement), as one may think. It is also a consequence of the hybrid divide-and-conquer setting; since the idea is to speed up the explorations of exponentially large trees by delegating subtrees of fractional height (say $\kappa' n$) to a quantum device, by construction, we still rely on the classical algorithm to explore the "top" (we denoted this tree $\mathcal{T}_0$ previously)

of the tree. This will generically take exponential time in the fraction $(1-\kappa')n$ which is still exponential in n.

In more detail, in the hybrid divide and conquer approach, the total run-time essentially attains the form $t_{hybrid} \in O^*(2^{\tau_h n})$, with $\tau_h = ((1-\kappa')\tau_c + \kappa'\tau_q)$ in the case of uniform trees where each (large enough) tree of height n' is of size $2^{\tau_c n'}$. We assume here that τ_q dictates the relative speed of the quantum algorithm (e.g. the quadratic speed-up of backtracking implies $\tau_q = \tau_c/2$. Even if the quantum query complexity and runtime was exactly zero (or, exponentially faster than the classical method), what remains is $O^*(2^{\tau_c(1-\kappa')n}))$. This constitutes a just polynomial speed up over $t_{classical} = O^*(2^{\tau_c n})$, which is the classical runtime on the entire tree.

This we summarize in the following lemma given without further proof.

Lemma 5.1. *The speed-up attainable by the hybrid divide and conquer approach with a κ'-effective size quantum computer is at best polynomial. The "speed-up" is subquadratic or quadratic at best, i.e. it holds that $t_{hybrid}(n) \in O((t_{classical}(n)))^{1-\alpha}$, where the degree of speed-up α is bounded $\alpha \leq \kappa'$ (if the quantum algorithm has polynomial run-time on the subtrees), and by $\alpha = 1/2$ (i.e. quadratic improvement) otherwise (if a Grover-type speed-up).*

The limitation of the above approach is that the quantum device gets used late in the game. Conceivably we can imagine settings where the quantum computer takes a more active role in the top of tree, or something similar. Indeed, beyond standard backtracking settings, better speed-ups are also possible, under mild, yet unavoidable assumptions.

In what follows we will assume that the evaluation of an arbitrary quantum circuit of size $\mathsf{poly}(n)$ on a classical computer takes exponential time. For concreteness we assume that solving BQP-complete promise problems, e.g. the problem of, for a given quantum circuit realizing some unitary U, determining whether the measurement outcome probability of one output qubit is 0 (of the register in the state $U|0\rangle^n$) is either larger than 2/3 or below 1/3 (under the promise that it is one of the two), requires $\Omega^*(2^{\tau n})$ classical computing steps (ignoring polynomial terms).[25] This is a circuit output problem.

In this case it is trivial to construct pathological examples of computational problems where exponential speed-ups can be attained (given a quantum computer of size $\kappa' n$) by arbitrarily choosing what the natural notion of the instance size should be. For instance, consider the circuit output problem, where the quantum circuit is special: no gates act on $(1-\kappa')n$ wires. In this case obviously a quantum computer of size $\kappa' n$ allows for a polynomial time solution, whereas the classical computer, by assumption requires $\Omega^*(2^{\kappa'\tau n})$ steps, which is an exponential separation for any κ'.

While this example is obviously pathological, one can easily imagine a more complicated yet related computation, where the n input bits are processed first by an involved classical computation which produces a specification of a quantum circuit on $\kappa' n$ wires, and the output of the overall computation is the output of that circuit.

This is an example of a broad spectrum of scenarios, where a (fewer-qubit) quantum computation is called as a subroutine of an over-arching classical computation.

One class of such computations are the hybrid approaches we investigate in this work. Another involves settings where classical computations are broken down, to distill the computationally hardest part, which is then delegated to a quantum machine, see e.g. [32].

It is also clear that, to obtain the best speed-ups, the quantum computer should be used at wherever possible in the computation, as is discussed in [33].

In what follows we consider a broad framework, the main purpose of which is to connect our work to less-than-obviously related works in quantum computing and quantum machine learning, such as the ones we exemplified above.

QuNets We wish to define a hybrid computational model which captures some of the facets of the limitations that quantum computers face in the near-term, namely size. We imagine access to a k-qubit quantum device, and want to consider all computations that can be run, when such a device is controlled, and augmented by,

[25] Unless we assume that quantum computations cannot be simulated in polynomial time, no better than polynomial improvements can be proved.

a (large) classical computer. This classical computer can pre-and-post process data, in between possibly many calls to the quantum device. In our model we will describe everything sequentially, although it will be a natural question how this can be parallelized when many k-sized quantum machines, which can communicate only classically, are available.

We name such a model a QuNet, and with $\mathrm{QuNet}(n,k)$ (with other qualifiers, described shortly) we denote the set of functions such a hybrid computation system can realize, given n-bit (classical) inputs and a k-qubit device. The number of output bits is specified when needed. It is worth noting that related models have been introduced and studied under various names by other authors, both explicitly and implicitly.[26] Specific to our setting, however, is the focus placed on the limitations of the qubit numbers k, relative to the instance size n.

It makes sense to distinguish two types of QuNets: adaptive (a-QuNet) and non-adaptive (QuNet). There are many ways to formalize both models, here we provide one approach; we define the latter first.

Let $\mathrm{cirqF}_k(\vec{c})$ denote the family of randomized functions, in which each randomized function takes a bitstring $\vec{c}$ as input and specifies the realization of a quantum circuit for some unitary U over k qubits. That is, the output of such a function is some k-sized bitstring $\vec{o}$, occuring with probability $|\langle\vec{o}|U|0^k\rangle|^2$, i.e. the function outputs what the quantum circuit would output.

$\mathrm{QuNet}(n,k)$ is then a (randomized) Boolean circuit, with n input wires, an arbitrary number of ancilla wires, where a standard Boolean gate-set is augmented with the gate set $\mathrm{cirqF}_k(\vec{c})$, which take $|\vec{c}|$ input classical wires,[27] and output k wires.

Note a $\mathrm{QuNet}(n,k)$ captures the two previous examples where an exponential speed up between a fully classical model ($\mathrm{QuNet}(n,0)$) and the genuine hybrid $\mathrm{QuNet}(n,\kappa' n)$ is provable, assuming quantum computers are not efficiently simulatable.

In the above model, the quantum computation is used essentially as a "black box". But in principle, more interaction is possible, once partial measurements are allowed. Here the classical computation can request that some of the quantum wires be measured, and the rest of the circuit may depend on the outcome.

The a-QuNet captures this additional freedom. It is easiest to characterize in a hybrid classical-quantum reversible circuit model common in quantum computing (where single wires are quantum, double classical) [34]. We consider a (quantum-like) circuit of n classical input wires, m other ancillary classical wires pre-set to zero, and a quantum register of k quantum wires pre-set in the state $|0\rangle$. We allow three types of gates: fully classical reversible gates (e.g. Toffoli and X – negation – will suffice); standard quantum gates, acting only on the k qubit wires; and CQ gates and QC gates. The CQ gates are classically controlled quantum gates, i.e. a quantum gate is applied depending on the state of a classical wire; QC gates are measurements: a number of quantum wires is measured in the computational basis, and the outcome is xored with the value of some target classical wires, matching in number. After measurement, we assume the state of that particular quantum wire is again re-set to $|0\rangle$.

It is not difficult to see that a-QuNets contain QuNets: the measurements are done on all wires, and where we note that any $\mathrm{cirqF}_k(\vec{c})$ gate can be implemented by using CQ gates. In terms of which functions they can realize, the two models are clearly equivalent; in fact, if complexity is not taken into account, the classical computation can simulate the entire quantum computation so indeed $\mathrm{QuNet}(n,0)$ is already universal. However in terms of efficiency and in other scenarios these two models can differ. For instance, a-QuNet captures error correction and fault-tolerant quantum computation protocols, and also the measurement-based quantum computing paradigm [36], where classical feedback from quantum measurements is extremely beneficial or assumed by construction.

Further it makes sense to limit the sizes of classical and quantum computations in both models, which allows for a more fine-grained comparison. With $\mathrm{QuNet}_{x,y}(n,k)$ we denote the func-

[26] For instance, the standard diagrammatic representation of circuits we can work with"double", classical wires, which can be classically processed, constitutes one such model [34, 35].

[27] There are many ways how a bitstring may specify a circuit, and how the circuit depth is encoded in the bitstring, but this is not relevant for us. All that matters is that some encoding exists.

tion family that can be realized using no more than x gates (including the quantum functions), and where the quantum circuits used use no more than y gates (note that in general $y \in O(|\vec{c}|)$). In the adaptive model we can simply count the classical gates vs the QC and CQ gates. Instead of particular values x and y can denote function families, e.g. poly or exp, so $x = \mathsf{poly}$ is a short hand for $x = O(\mathsf{poly}(n))$.[28]

With complexity-theoretical considerations, the relationship between a-QuNet(n, k) and QuNet(n, k) is not entirely clear, in that, in general, classical adaptivity may reduce the number of quantum gates needed – it is well known that any classical control can be raised to fully quantum (with no classical feed-back), but at the expense of more quantum wires. For instance in the case of an algorithm using QPE to some precision ϵ, in many cases it is known that one can perform most computations using 1 ancillary wire adaptively[29] (instead of $\log(1/\epsilon)$ ancillas which can be measured at once). In turn it is not clear the same can be done non-adaptively, where at each step the entire register must be measured, without at least $\mathsf{polylog}(1/\epsilon)$ multiplicative additional computational costs (see [37] for state-of-art approaches to single qubit QPE).

More generally, to our knowledge, not much is known about the costs of rewriting an adaptive circuit with partial measurements as a circuit with complete measurements without introducing ancillary qubits; but efficient methods for this could simplify the execution of quantum algorithms on small machines that are also limited in coherence times. This topic goes beyond the scope of the present paper.

We finalize this section by highlighting the connections between our hybrid model and other related lines of investigation, in the context of QuNets.

One example is our hybrid divide and conquer method, where the corresponding QuNet(n, k) has a near-trivial repeating structure as all quantum computations are of the same type, tackling smaller problem instances generated by a classical pre-processing step (the 'top of the tree'). In particular, previous results in the hybrid approach are examples of $\mathrm{QuNet}_{\mathsf{exp},\mathsf{exp}}(n, k)$ which solve various NP-hard problems exactly, faster than their classical counterparts. In the context of quantum annealers, all schemes developed for the purpose of fitting a larger computation on a fixed sized device (e.g. see [38]) fit in this paradigm, and they ostensibly 'get more' out of the device, however mostly in a heuristic setting where little can be proved. These are examples of $\mathrm{QuNet}_{\mathsf{poly},\mathsf{poly}}(n, k)$ which tackle various NP-hard and quantum chemistry problems heuristically. A relatively recent paper that also focuses on getting the most out of a smaller device utilizes data reduction, see [33]. In all these examples, the computational problem is from a classical domain, and the approach is to 'quantize' subroutines.

In an opposite vein, in [39, 40], the authors present hybrid computations which compute the output of a large quantum circuit (given on input), calling a smaller quantum device. These are examples of a $\mathrm{QuNet}_{\mathsf{exp},\mathsf{poly}}(n, k)$ which solves the problem of simulating quantum computations. In this case, only the number of calls to the quantum device, and the classical processing is exponential, whereas the quantum circuits are polynomially sized. It is clear that constructing QuNets with small k for some hard problem is appealing for near-term quantum devices. However, it is not clear what are the families of functions and the complexity classes which can be captured by this model.

Similarly the relationship between QuNets and classical and quantum parallel complexity classes which also care about splitting of the computation on multiple units (however, not caring about the required space) also remains to be clarified. Coming up with a structure and a QuNet algorithm for a target problem may be difficult.

In the domain of variational methods, specifically applied to machine learning this problem could be circumvented. Any such network where the quantum circuits are externally parametrized is a valid Ansatz for a parametrized approach to supervised learning, or to generative modeling. In particular, such networks generalize neural networks. Individual neurons are replaced by a circuit, a subset of parameters of which depend on the input values, and the remainder is free. The free parameters play the role of tunable weights in an artificial neuron. Ways

[28] Note that with these limitations the number of classical ancillas we need to allow is also bounded by $x + y$.

[29] All that is required is that the ancillary qubit is measured, and reset for the next step in QPE.

to construct meaningful QuNets for machine learning purposes of this type are a matter of ongoing research.

Acknowledgements

VD thanks Tom O'Brien for discussions regarding adaptive and non-adaptive QuNets. This research was supported by the Dutch Research Council (NWO/OCW), as part of the Quantum Software Consortium program (project number 024.003.037), and has been sponsored by the European Union's Horizon 2020 research and innovation program under the NEASQC project, grant agreement No 951821.

References

[1] Vedran Dunjko, Yimin Ge, and J Ignacio Cirac. "Computational speedups using small quantum devices". Physical review letters **121**, 250501 (2018).

[2] Yimin Ge and Vedran Dunjko. "A hybrid algorithm framework for small quantum computers with application to finding hamiltonian cycles". Journal of Mathematical Physics **61**, 012201 (2020).

[3] Ashley Montanaro. "Quantum-Walk Speedup of Backtracking Algorithms". Theory of Computing **14**, 1–24 (2018).

[4] Martin Davis, George Logemann, and Donald W Loveland. "A machine program for theorem-proving". New York University, Institute of Mathematical Sciences. (1961).

[5] Ramamohan Paturi, Pavel Pudlák, Michael E Saks, and Francis Zane. "An improved exponential-time algorithm for k-SAT". Journal of the ACM (JACM) **52**, 337–364 (2005).

[6] Earl Campbell, Ankur Khurana, and Ashley Montanaro. "Applying quantum algorithms to constraint satisfaction problems". Quantum **3**, 167 (2019).

[7] Simon Martiel and Maxime Remaud. "Practical implementation of a quantum backtracking algorithm". In International Conference on Current Trends in Theory and Practice of Informatics. Pages 597–606. Springer (2020).

[8] Thomas Dueholm Hansen, Haim Kaplan, Or Zamir, and Uri Zwick. "Faster k-sat algorithms using biased-ppsz". In Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing. Pages 578–589. STOC 2019New York, NY, USA (2019). ACM.

[9] Armin Biere, Marijn Heule, and Hans van Maaren. "Handbook of satisfiability". Volume 185. IOS press. (2009). url: https://dl.acm.org/doi/10.5555/1550723.

[10] Marc Mezard, Marc Mezard, and Andrea Montanari. "Information, physics, and computation". Oxford University Press. (2009).

[11] Martin Davis and Hilary Putnam. "A Computing Procedure for Quantification Theory". J. ACM **7**, 201–215 (1960).

[12] Marijn JH Heule, Matti Juhani Järvisalo, Martin Suda, et al. "Solver and benchmark descriptions". In Proceedings of SAT competition 2018. Department of Computer Science, University of Helsinki (2018). url: http://hdl.handle.net/10138/237063.

[13] Robert Nieuwenhuis, Albert Oliveras, and Cesare Tinelli. "Abstract DPLL and abstract DPLL modulo theories". In International Conference on Logic for Programming Artificial Intelligence and Reasoning. Pages 36–50. Springer (2005).

[14] Jasmin Christian Blanchette, Sascha Böhme, and Lawrence C Paulson. "Extending Sledgehammer with SMT solvers". Journal of automated reasoning **51**, 109–128 (2013).

[15] Daniel Rolf. "Derandomization of PPSZ for unique-k-SAT". In International Conference on Theory and Applications of Satisfiability Testing. Pages 216–225. Springer (2005).

[16] Dominik Scheder and John P Steinberger. "PPSZ for general k-SAT-making Hertli's analysis simpler and 3-SAT faster". In 32nd Computational Complexity Conference (CCC 2017). Volume 79, pages 9:1–9:15. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik (2017).

[17] Daniel Rolf. "Improved bound for the PPSZ/schoning-algorithm for 3-SAT". Journal on Satisfiability, Boolean Modeling and Computation **1**, 111–122 (2006).

[18] Timon Hertli. "3-SAT faster and simpler—

unique-SAT bounds for PPSZ hold in general". SIAM Journal on Computing **43**, 718–729 (2014).

[19] Timon Hertli. "Breaking the PPSZ barrier for unique 3-SAT". In Automata, Languages, and Programming: 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part I 41. Pages 600–611. Springer (2014).

[20] Dominik Scheder. "PPSZ is better than you think". In 2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS). Pages 205–216. (2022).

[21] Lov K Grover. "A fast quantum mechanical algorithm for database search". In Proceedings of the twenty-eighth annual ACM symposium on Theory of computing. Pages 212–219. (1996).

[22] Andris Ambainis. "Quantum search algorithms". ACM SIGACT News **35**, 22–35 (2004).

[23] Andris Ambainis and Martins Kokainis. "Quantum algorithm for tree size estimation, with applications to backtracking and 2-player games". In Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing. Pages 989–1002. (2017).

[24] Michael Jarret and Kianna Wan. "Improved quantum backtracking algorithms using effective resistance estimates". Physical Review A **97**, 022337 (2018).

[25] Dominic J. Moylett, Noah Linden, and Ashley Montanaro. "Quantum speedup of the traveling-salesman problem for bounded-degree graphs". Phys. Rev. A **95**, 032323 (2017).

[26] Neil D. Jones and William T. Laaser. "Complete problems for deterministic polynomial time". Theoretical Computer Science **3**, 105 – 117 (1976).

[27] Raymond Greenlaw, H James Hoover, Walter L Ruzzo, et al. "Limits to parallel computation: P-completeness theory". Oxford University Press on Demand. (1995). url: https://dl.acm.org/doi/book/10.5555/203244.

[28] Thomas Lengauer and Robert E Tarjan. "Asymptotically tight bounds on time-space trade-offs in a pebble game". Journal of the ACM (JACM) **29**, 1087–1130 (1982).

[29] Charles H Bennett. "Time/space trade-offs for reversible computation". SIAM Journal on Computing **18**, 766–776 (1989).

[30] T. Altman and Y. Igarashi. "Roughly sorting: Sequential and parallel approach". J. Inf. Process. **12**, 154–158 (1989). url: http://id.nii.ac.jp/1001/00059782/.

[31] Hong-Yu Liang and Jing He. "Satisfiability with index dependency". Journal of Computer Science and Technology **27**, 668–677 (2012).

[32] Marcello Benedetti, John Realpe-Gómez, and Alejandro Perdomo-Ortiz. "Quantum-assisted Helmholtz machines: A quantum–classical deep learning framework for industrial datasets in near-term devices". Quantum Science and Technology **3**, 034007 (2018).

[33] Aram W. Harrow. "Small quantum computers and large classical data sets" (2020) arXiv:2004.00026.

[34] Michael A Nielsen and Isaac Chuang. "Quantum computation and quantum information". American Association of Physics Teachers. (2002).

[35] Robert B. Griffiths and Chi-Sheng Niu. "Semiclassical fourier transform for quantum computation". Phys. Rev. Lett. **76**, 3228–3231 (1996).

[36] Robert Raussendorf and Hans J. Briegel. "A one-way quantum computer". Phys. Rev. Lett. **86**, 5188–5191 (2001).

[37] Thomas E O'Brien, Brian Tarasinski, and Barbara M Terhal. "Quantum phase estimation of multiple eigenvalues for small-scale (noisy) experiments". New Journal of Physics **21**, 023022 (2019).

[38] Akshay Ajagekar, Travis Humble, and Fengqi You. "Quantum computing based hybrid solution strategies for large-scale discrete-continuous optimization problems". Computers & Chemical Engineering **132**, 106630 (2020).

[39] Tianyi Peng, Aram W Harrow, Maris Ozols, and Xiaodi Wu. "Simulating large quantum circuits on a small quantum computer". Physical review letters **125**, 150504 (2020).

[40] Sergey Bravyi, Graeme Smith, and John A. Smolin. "Trading classical and quantum computational resources". Physical Review X**6** (2016).

[41] Martijn Swenne. "Solving SAT on noisy quantum computers". https://theses.liacs.nl/1725 (2019). Bachelor thesis.

[42] Theodore J Yoder, Guang Hao Low, and Isaac L Chuang. "Fixed-point quantum search with an optimal number of queries". Physical review letters**113** (2014).

[43] Alessandro Cosentino, Robin Kothari, and Adam Paetznick. "Dequantizing Read-once Quantum Formulas". In 8th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2013). Volume 22 of Leibniz International Proceedings in Informatics (LIPIcs), pages 80–92. Dagstuhl, Germany (2013). Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[44] David A Barrington. "Bounded-width polynomial-size branching programs recognize exactly those languages in NC1". Journal of Computer and System Sciences **38**, 150–164 (1989).

A Proof of Proposition 2.1

In this section, we prove that dncPPSZ (see Algorithm 2) is correct and has a runtime of $2^{(\gamma_k+\varepsilon)n}$, i.e., it equals the best-known SAT algorithm as recent analysis of PPSZ confirms [20] (it surpassed Biased PPSZ [8] again). We focus here on unique-k-SAT, in which the input formula has a unique satisfying assignment if satisfiable. (The results can be generalized to general k-SAT based on [20].) Let F be a k-SAT formula on n variables, $\vec{u}$ be the unique satisfying assignment of F (provided F is satisfiable).

Proving the runtime of dncPPSZ is easy, as shown by Theorem A.1, which we mainly introduce to define the best-known value of γ_k.

Theorem A.1. *Algorithm 2 has a runtime of $2^{(\gamma_k+\varepsilon)n}$ with $\gamma_3 \approx 0.386229$ [20].*

Proof. The algorithm only explores a tree truncated to depth $(\gamma_k+\varepsilon)n$ for a given (fixed) permutation π. □

We now show that dncPPSZ is a correct Monte-Carlo algorithm, i.e., it finds the unique satisfying assignment with constant probability if it exists (hence, if repeated a constant number of times, it will decide unique satisfiability). To this end, we first reiterate some results from the PPSZ algorithm studied in [20].

In contrast to the dncPPSZ algorithm, which tries all assignments with backtracking, recent versions of PPSZ [20] sample assignments using an *advice string* $\vec{a}$ and the *Decode* function shown in Algorithm 4. According to a random variable order π, the function only uses a bit of the advice string $\vec{a}$ when the next variable is guessed, i.e., when the heuristic fails to establish its truth value (see the `else` branch). It also does not terminate evaluation early, as in backtracking, i.e., it naively continues evaluation even when the formula is already (un)satisfied. This suffices because the only point of the Decode function is to study the expected number of guesses under s-implication and to demonstrate how the number of guesses dictates the expectation of finding the satisfying assignment.

Theorem A.2 shows that Decode uses $\gamma_k n$ guesses to find the unique satisfying assignment of a k-SAT formula F. By Corollary A.3, it can be found with high probability with a subexponential number of (random) calls to Decode.

Algorithm 4 Decode(F, $\vec{a}$, π) where the value s of s-implication is a large constant or grows very slowly in n [20]. The function returns false or the unique satisfying assignment, as a CNF formula, if the advice $\vec{a}$ and order π 'decode' to it.

```
if π = ⟨⟩ then return F = ∅                          ▷ SAT?
x ← π[0]                                              ▷ first var in π
if F ⊨_s (x = b) for b ∈ {0,1} then                   ▷ forced:
    c := b
else                                                  ▷ guessed:
    c := a[0]
    a⃗ := a⃗[1...]                                      ▷ postfix of array
return Decode(F_{|x=c}, a⃗, π[1...]) ∧ (x = c)
```

The PPSZ algorithm is therefore defined as the loop that calls Decode sufficiently many times.

Theorem A.2 (Adapted from [20] Th.4)**.** *Let formula F, advice $\vec{a}$ and order π be such that Decode$(F,\vec{a},\pi) = \vec{u}$ (Algorithm 4). Then the expected number of guesses in Decode equals $\gamma_k n$.*

Corollary A.3 ([20] Ob.5)**.** *The unique satisfying assignment is found with high probability with $O^*(2^{(\gamma_k+\varepsilon)n})$ calls to Decode with random assignment $\vec{a}$ and order π.*

Instead of trying random advices, the dncPPSZ algorithm searches all advices of length $(\gamma_k+\varepsilon)n$. Since it implements the Decode function along each search branch, merely adding backtracking when a formula becomes (un)satisfiable on a partial assignment, it also inherits the probability of finding the unique satisfying assignment from PPSZ. Theorem A.4 shows this. (It implies Proposition 2.1.)

Theorem A.4. *Let F be a k-SAT formula defined on n variables and π a permutation in S_n. Then the probability that dncPPSZ$_\pi$ returns a satisfying assignment is at least $1 - \frac{1}{1+\varepsilon/\gamma_k}$.*

Proof. Let $g(\pi)$ denote the number of guesses Decode performs for $\vec{a}$ with Decode$(F,\vec{a},\pi) = \vec{u}$. By applying Markov's inequality, we obtain:

$$\begin{aligned}\Pr_\pi\left[g(\pi) > (\gamma_k+\varepsilon)n\right] &\le \Pr_\pi\left[g(\pi) \ge (\gamma_k+\varepsilon)n\right] \\ &\le \frac{\mathbb{E}_\pi[g(\pi)]}{(\gamma_k+\varepsilon)n} = \frac{\gamma_k}{\gamma_k+\varepsilon},\end{aligned}$$

which is an upper bound on the probability that dncPPSZ$_\pi$ fails to find a satisfying assignment, given it exists. Subsequently, $1 - \frac{\gamma_k}{\gamma_k+\varepsilon}$ is then a lower bound on the probability that dncPPSZ$_\pi$ succeeds. □

B Space-efficient quantum subroutines for tree search algorithms

In this section, we implement quantum subroutines for tree search algorithms. We implement a space-efficient Grover-based search for k-SAT, and then explain how to efficiently implement Quantum Phase Estimation (QPE) for realizing quantum backtracking.

B.1 Space-efficient formula evaluation oracles for k-SAT

To implement Grover search of Figure 3 for k-SAT, the oracle $\mathcal{O}_f$ should evaluate a formula F on an assignment of n variables. A naive oracle implementation uses $n+m+1$ qubits for a formula of m clauses, including $p \equiv m+1$ ancilla qubits to evaluate each clause and their conjunction. This oracle implementation evaluates each clause and stores the result in a dedicated ancilla qubit. Figure 2 shows an example which uses a 3-controlled Toffoli gate. After all clauses are evaluated, an m-controlled Toffoli gate can computed the final result from the dedicated clause qubits. The reversal is trivial.

More efficient oracle implementations are possible: previous work [41] has shown that multi-controlled incrementers can be used to reduce the ancilla count p from $m+1$ to $\lceil \log(m) \rceil + 1$. Furthermore, using Fixed-Point Amplitude Amplification [42], the ancilla count can be further reduced to 2 [41].

In the present work, we exploit another space-efficient quantum algorithm for formula evaluation, which only uses 1 ancilla qubit, relying on prior work [43] which defines an one-qubit program which computes any Boolean function exactly: this program takes x as input and the output of the measurement is $f(x)$. This one-qubit model, which alternates single-qubit unitaries and controlled gates, can be seen as a quantum variant of Barrington's theorem [44]. Note

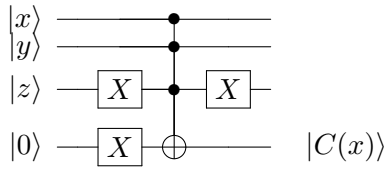

Figure 2: Naïve implementation of the evaluation of clause $C \equiv (\neg x \vee \neg y \vee z) \equiv \neg(x \wedge y \wedge \neg z)$

that k-CNF formulas (which are CNF formulas) correspond to depth-2 (AC) circuits, so the theorem (a statement on NC_1) applies to k-SAT input.

We can use the one-qubit model to implement the oracle $\mathcal{O}_f$, using one ancilla qubit to apply single-qubit unitaries and CNOT gates, controlled over the qubits of the input. We refer the interested reader to [43, Sec. 4] for the details of the implementations of the single-qubit unitaries of the one-qubit model.

Using the one-qubit model for formula evaluation to implement the oracle, we obtain a space-efficient implementation of Grover which uses 1 qubit for each variable of the formula, 1 qubit for the phase-flip oracle, and 1 qubit for formula evaluation, for a total space complexity of $n+2$, as shown in Figure 3 with $p=1$.

B.2 Space-efficient quantum phase estimation

The quantum backtracking framework which is used in this paper relies on the Quantum Phase Estimation (QPE) algorithm, a quantum algorithm which outputs with high probability a t-bit estimate of the phase of its inputs: a unitary U and an eigenvector $|\psi\rangle$. The canonical implementation of QPE uses t qubits to produce its t-bit estimate. In this section, we study a more space efficient implementation of a specific version of QPE, sufficient for quantum backtracking, which only determines whether the phase is 0.

Proposition B.1. *There is an implementation of QPE which checks whether the phase of the eigenvector (of a given unitary U operating on m qubits) is equal to zero with a t-bit precision using no more than $p+1$ ancilla qubits, where $p = \lceil \log(t) \rceil$ qubits are allocated for a counter from 0 to t.*

Proof. Consider the original QPE circuit represented in Figure 4, which uses t ancilla qubits to obtain a t-bit estimate of θ such that $U|\psi\rangle = e^{2i\pi\theta}|\psi\rangle$, given $|\psi\rangle$ as input.

From the initial state $|0\rangle^{\otimes t}|\psi\rangle$, we apply the Hadamard gate H to each qubit of the first register, obtaining the state $|+\rangle^{\otimes t}|\psi\rangle$. We then apply unitaries $U, \ldots, U^{2^{t-1}}$, each controlled over one qubit of the first register, obtaining the state

$$|\psi'\rangle = \frac{1}{\sqrt{2^t}} \bigotimes_j \left(|0\rangle + e^{2i\pi 2^j \theta}|1\rangle\right)$$

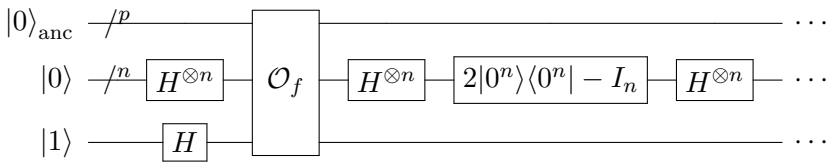

Figure 3: Grover's algorithm: the oracle $\mathcal{O}_f$ xor's the result on the phase wire (bottom).

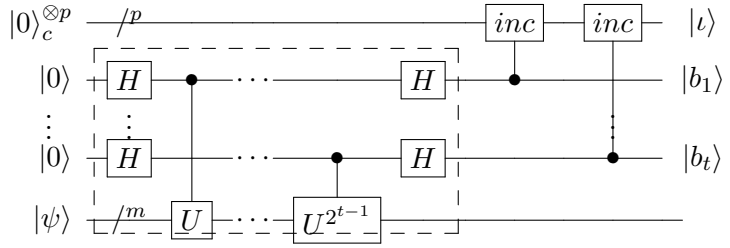

Figure 4: The specialized QPE circuit, which which also counts the bits of the phase that are equal to 1. The circuit in the dotted rectangle determines whether the t-bit estimate of θ is null (as in canonical QPE).

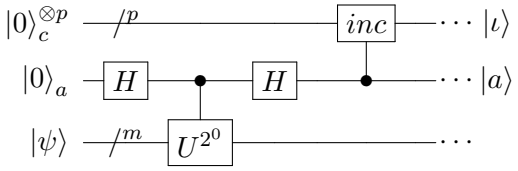

Figure 5: Space efficient circuit for determining whether the t-bit estimate of θ is null. The circuit is repeated for $U^{2^1}, \ldots, U^{2^{t-1}}$

To this state, we apply $H^{\otimes t}$, obtaining the final state.

$$|\psi_{\text{f}}\rangle = \frac{1}{2^t} \bigotimes_j \left((1 + e^{2i\pi 2^j \theta})|0\rangle + (1 - e^{2i\pi 2^j \theta})|1\rangle\right)$$

Let us write p_0 for the probability that the t-bit estimate that in the final state $|\psi_{\text{f}}\rangle$ of this circuit is $b_1 \ldots b_t = 0 \ldots 0$, so that $p_0 = |\alpha_0|^2$ where

$$\alpha_0 = \frac{1}{2^t} \prod_j \left(1 + e^{2i\pi 2^j \theta}\right)$$

Now, consider the circuit defined in Figure 5. It is a more space-efficient implementation of QPE which only detects whether the phase is equal to 0. It can be implemented using $\lceil \log(t) \rceil$ ancilla qubits, which add up to the number of qubits on which U is defined.

This implementation stems from the following observation. Consider the QPE circuit which detects whether the phase is equal to 0, to which we add an ancilla register which counts up to t (requireing $p = \lceil \log(t) \rceil$ qubits) and gates which increment the counter whenever one of the bits of the t-bit estimate is not equal to 0, as depicted in Figure 4. Since the counter is only increased (and never decreased) throughout the computation, the value of the counter is only non-zero in the final state if a 1 appeared in one of the bits. In other words, the counter can only have a null value when the all zero state is the t-bit estimate.

This observation leads us to the implementation of a circuit which, for each j ($0 \leq j < t$), applies the Hadamard gate H on an ancilla qubit a, followed by the unitary U^{2^j} controlled on a, the Hadamard gate H on a, and an incrementation of the counter if the value of a is non-zero, as pictured in Figure 5.

For any j, assume that the counter is in the all zero state and a is also $|0\rangle$. From our previous observation, this means that the counter has not been increased yet. Write $|\psi_j\rangle$ and $\left|\psi'_j\right\rangle$ respectively for the overall states before and after the controlled incrementation.

$$|\psi_0\rangle = |0\rangle_c^{\otimes p} \frac{1}{2} \left((1 + e^{2i\pi\theta})|0\rangle + (1 - e^{2i\pi\theta})|1\rangle\right)$$

$$\begin{aligned} |\psi'_0\rangle = {} & \frac{1}{2}|0\rangle_c^{\otimes p}(1 + e^{2i\pi\theta})|0\rangle \\ & + \frac{1}{2}(1 - e^{2i\pi\theta})|0 \ldots 01\rangle_c|1\rangle \end{aligned}$$

The state $|\psi'_0\rangle$ contains a 'non-zero branch', which carries over throughout the next step of the computation, so that

$$\begin{aligned} |\psi_1\rangle = {} & \frac{1}{2}(1 + e^{2i\pi\theta})|0\rangle_c^{\otimes p} \\ & \otimes \frac{1}{2}\left((1 + e^{4i\pi\theta})|0\rangle + (1 - e^{4i\pi\theta})|1\rangle\right) \\ & + \text{ non-zero branch} \end{aligned}$$

$$|\psi_1'\rangle = \frac{1}{4}(1+e^{2i\pi\theta})(1+e^{4i\pi\theta})|0\rangle_c^{\otimes p}|0\rangle$$
$$+ \text{ non-zero branch}$$
$$\vdots$$
$$|\psi_t'\rangle = \frac{1}{2^t}\prod_{j=0}^{t-1}(1+e^{2\pi i 2^j\theta})|0\rangle_c^{\otimes p}|0\rangle$$
$$+ \text{ non-zero branch}$$

where $|\psi_t'\rangle$ is the final state $|\psi_{\mathrm{f}}'\rangle$ of the circuit.

The probability p_0' of obtaining a zero value in the final state of the counter is $p_0' = |\alpha_0'|$ where

$$\alpha_0' = \frac{1}{2^t}\prod_j \left(1+e^{2i\pi 2^j\theta}\right) = \alpha_0$$

Therefore, $p_0 = p_0'$, which means that the probability of observing the all zero t-bit estimate in the first circuit and the probability of observing the value 0 in the counter of the second circuit are exactly the same. By linearity this observation holds for all input states, implying that we have come up with a more efficient implementation of QPE which detects whether the phase is equal to 0 (up to a t-bit estimate) while using only $\lceil \log(t) \rceil$ qubits. □

Note that this construction is important to the present work, as the quantum backtracking framework's use of QPE requires a $O(\log(T))$-bit precision, where T is the size of the search tree considered. When T is exponential in n, the standard implementation of QPE yields a linear overhead which does not prevent us from using the hybrid approach, but directly weakens the computational efficiency of hybrid schemes. Bringing down QPE's overhead from $O(\log(T))$ to $O(\log(\log(T)))$ implies almost no loss between real and effective quantum computer size for many problems.

C Quantum backtracking for DPLL-like algorithms

In this section, we develop a space-efficient implementation of the quantum backtracking framework [3] for DPLL-like algorithms. It is an essential element for the hybrid approach developed throughout this paper: Whenever the classical computation hands of a sub-problem corresponding to a (restricted) formula F to the quantum computer, it generates the space-efficient quantum circuit from F as described here.

As DPLL provides no guarantee on the maximal number of guesses which have to be done to reach a satisfying assignment, we represent tree vertices with the full partial assignment they correspond to, rather than the branching choices (guesses) as we do for PPSZ in Appendix E.

C.1 The quantum backtracking framework

We present the quantum backtracking framework developed in [3], closely following their exposition.

The search tree underlying a classical k-SAT-solving backtracking algorithm is formalised as a rooted tree $\mathcal{T}$ of depth n and with T vertices $r, 1, \ldots, T-1$. Each vertex is labeled by a partial assignment, and a marked vertex is a vertex labeled by a satisfying assignment. We work under the promise that the formula given as input is not trivially satisfied, and therefore the root is promised not to be marked.

We write $\ell(x)$ for the distance from the root r to a vertex x, and assume that $\ell(x)$ can be determined for each vertex x (even without full knowledge of the structure of the tree). The examples developed in this paper make this consideration trivial, as vertices are labeled by list of variable assignments, with the special symbol $*$ marking unassigned variables, and therefore the distance from the root can be calculated from the number of assigned variables.

We write A (resp. B) for the set of vertices an even (resp. odd) distance from the root, with $r \in A$. We write $x \to y$ to mean that y is a child of x in the tree. For each x, let d_x be the degree of x as a vertex in an undirected graph. So for every vertex x which isn't the root, we have $d_x = |\{y \mid x \to y\}| + 1$ and $d_r = |\{y \mid r \to y\}|$.

We define the quantum walk[30] as a set of diffusion operators D_x on the Hilbert space $\mathcal{H}$ spanned by $\{|r\rangle\} \cup \{|x\rangle : x \in \{1, \ldots, T-1\}\}$, where D_x acts on the subspace $\mathcal{H}_x$ spanned by $\{|x\rangle\} \cup \{|y\rangle : x \to y\}$. We take $|r\rangle$ to be its initial state.

Such diffusion operators D_x are defined as the identity if x is marked, and as follows otherwise:

[30] Note that this notion of quantum walk does not involve a separate "coin" space.

- for $x \neq r$, $D_x = I - 2|\psi_x\rangle\langle\psi_x|$, where

$$|\psi_x\rangle = \frac{1}{\sqrt{d_x}}\left(|x\rangle + \sum_{y,x\to y}|y\rangle\right).$$

- $D_r = I - 2|\psi_r\rangle\langle\psi_r|$, where

$$|\psi_r\rangle = \frac{1}{\sqrt{1+d_r n}}\left(|r\rangle + \sqrt{n}\sum_{y,r\to y}|y\rangle\right).$$

Note that when x is an unmarked leaf (in the context of k-SAT, a vertex which corresponds to a contradiction), it has no neighbors and therefore the reflectors are about the state itself.

We define two Szegedy-style walk operators as follows:

$$R_A = \bigoplus_{x\in A} D_x \text{ and } R_B = |r\rangle\langle r| + \bigoplus_{x\in B} D_x.$$

Assume you have access to R_A and R_B, and consider the following algorithm.

Detecting a marked vertex
Input: Operators R_A, R_B, a failure probability δ, upper bounds on the depth n and the number of vertices T. Let $\beta, \gamma > 0$ constants given in Ashely's paper.

(a) Repeat the following subroutine $K = \lceil \gamma \log(1/\delta) \rceil$ times:

i. Apply phase estimation to the operator $R_B R_A$ with precision $\beta/\sqrt{Tn}$, on the *initial* state $|r\rangle$.

ii. If the eigenvalue is 1, accept; otherwise, reject.

(b) If the number of acceptances is at least $3K/8$, return "marked vertex exists"; otherwise, return "no marked vertex".

In essence, this algorithm detects whether the tree T has a marked vertex using $O(\sqrt{Tn}\log(1/\delta))$ queries to R_A, R_B. Detection is enough: to find a satisfying assignment, it suffices to traverse the tree and ask which subtree leads to a satisfying assignment at each branching. The remainder of this appendix is dedicated to a space-efficient implementation of quantum backtracking for DPLL-like algorithms.

C.2 Encoding sets of variables

We first describe how partial assignments are encoded and define subroutines which allow us to manipulate assignments. As a partial assignment for a set Vars of variables can be seen as a function Vars $\to \{0, 1, *\}$, a vertex x can be uniquely labeled in this n-trit system, and can hence be stored in a quantum state $|\vec{x}\rangle$ of $\log_2(3)n$ qubits.

Since we have the restricted formula available whenever the classical algorithm constructs a quantum circuit, we could also hard code the wires according to the clauses, but to ease drawing of the circuits, we use a unitary instead. Whether a given variable x_i is in a clause C_j can be easily checked using a Toffoli gate and Pauli-X gates to determine whether a given index i is in the clause or not. Such unitaries form a family of unitaries

$$\text{Check}_{j,i} : |\vec{x}\rangle|0\rangle|0\rangle \mapsto |\vec{x}\rangle|b\rangle|s\rangle,$$

where $b = 1, s = 0$ ($s = 1$) if the literal x_i ($\overline{x_i}$) appears in C_j.

To realize the search predicate $P(\vec{x})$, i.e, whether $\vec{x}$ is a leaf of the search tree, we define the unitary V_{leaf}. This is a unitary such that $V_{\text{leaf}}|\vec{x}\rangle|0\rangle = |\vec{x}\rangle|b\rangle$, where b is a Boolean value equal to 1 if and only if $\vec{x}$ corresponds to a trivial formula (whenever $P(\vec{x})$ should be either 0 or 1, as the formula $F_{|\vec{x}}$ is (un)sat). The find solutions, we make use of the unitary V_{marked} such that $V_{\text{marked}}|\vec{x}\rangle|0\rangle = |\vec{x}\rangle|b\rangle$, where b is a Boolean value equal to 1 if and only if $\vec{x}$ is a satisfying assignment. Those unitaries are hardcoded based on the restricted formula, i.e, by creating a reversible circuit for $F_{|\vec{x}}$.

C.3 Implementing the walk operator

We describe step by step an implementation of the walk operator $W = R_B R_A$, for DPLL-like backtracking algorithms. Let $\vec{x}$ be a partial assignment of the formula (i.e. a vertex in our search tree). We provide a quantum reversible implementation of the routines $ch1(\vec{x}), ch2(\vec{x}, b)$ and $chNo(\vec{x})$ specified in Section 2.4.

We define reversible routines which implement both the reduction rule and the branching heuristic (see Section 2.4), in order to check whether the unit rule or the pure literal rule can be applied to one of the unassigned variables in assignment $\vec{x}$ according to a fixed (static) variable

ordering. In other words, the first unassigned variable x for which there is a clause $C \in F_{|\vec{x}}$ that is unit, i.e., $C = \{x\}$ or $C = \{\bar{x}\}$, is forced accordingly. The same is done for the pure literal rule. We provide implementations of the unit rule in Appendix C.4 and of the pure literal rule in Appendix C.5.

We implement $ch1(\vec{x}), ch2(\vec{x}, b)$ and $chNo(\vec{x})$ as unitaries V_1 and V_2 which respectively compute the first child $\vec{x}_1$ (assuming $\vec{x}$ is a non-leaf), and the second child $\vec{x}_2$, assuming $\vec{x}$ is not forced (guessed) and non-leaf. In other words, in the 2 children case, V_i implements $\vec{x} \to ch2(\vec{x}, i)$. The specification adds V_0 for computing the identity function for leaf $|\vec{x}\rangle$ i.e., $\vec{x}_0 \triangleq \vec{x}$:

$$V_i : |\vec{x}\rangle\left|\vec{0}\right\rangle \mapsto |\vec{x}\rangle|\vec{x}_i\rangle$$

We implement a unitary $V_A(c)$ which, given a vertex $\vec{x}$ outputs the superposition of the vertex $\vec{x}$ and its c children. Its specification is:

$$\begin{aligned} V_A(c) : |\vec{x}\rangle\left|\vec{0}\right\rangle|0\rangle &\xmapsto{H} \frac{1}{\sqrt{c}} \sum_{0\le i\le c} |\vec{x}\rangle\left|\vec{0}\right\rangle|i\rangle \\ &\xmapsto{\text{ctrl-}V_i} \frac{1}{\sqrt{c}}|\vec{x}\rangle \sum_{0\le i\le c} |\vec{x}_i\rangle|i\rangle \\ &\xmapsto{V_C(c)} \frac{1}{\sqrt{c}}|\vec{x}\rangle \sum_{0\le i\le c} |\vec{x}_i\rangle|0\rangle \end{aligned}$$

In other words, we apply each ctrl-V_i controlled over a qutrit (the 'index register') which determines which V_i to apply, with $i = 0$ being the operation which copies the parent vertex. The result is an entangled state between the index register and specification of the children, rather than a superposition over children. The operation $V_C(c)$ disentangles the state, using the fact that we can prepare, and hence, unprepare the i-th child. For each child index i, we erase the index register as follows; we uncompute the i-th child by inverting V_i. Now, in the child register, in the branch with the i-th child we have the "all-zero" state. Since all children differ, this is the only branch with the all-zero state. Then, conditioned on the child-specifying register being in the all-zero state, we subtract i, from the child-index register. Finally, we recompute the i-th child, which restores the children specifications in all branches. This erases the which-child information for each child, leaving the index register in the $|0\rangle$ state for all children.

Now, the operator R_A can be implemented as $U_A\left(\mathrm{Id} - |0\rangle\langle 0|\right)U_A^\dagger$, where $U_A = \bigoplus_{\vec{x}\in A} U_A^{\vec{x}}$ is the unitary which computes the state $|\varphi_x\rangle$, i.e.,

$$U_A|\vec{x}\rangle|0\rangle = |\vec{x}\rangle|\varphi_{\vec{x}}\rangle$$

Recall that the implementation of the diffusion operators depends on the parity of their distance from the root. We implement each D_x for $x \in A$ as a unitary U_A^x by checking whether $\vec{x}$ has zero, one or two children. This is done by checking the number c of children, and controlled on the result bit of this operation, applying (or not) the corresponding operation $V_A(c)$ which generates the superposition over the child(ren) and original vertex.

The operator R_B is implemented in a similar fashion to R_A, assuming that we have access to a unitary V_{root} which checks whether $\vec{x} = r$. Observing that the root r is associated to the all-undetermined satisfying assignment, i.e. where $\mathbf{r} = *^n$. Such a unitary can easily be implemented with a counter to n and incrementation controlled on each variable being equal to $*$.

In the generic quantum backtracking algorithm, a depth counter is maintained to determine the parity of the depth at which the vertex $\vec{x}$ is. We forgo of such a register by observing that each variable assignment takes us one level deeper into the tree, and therefore the depth at which $\vec{x}$ is at is given is defined by the number of variables which have already been assigned a value. Therefore, to check the parity of the depth, it suffices to count every time $x_i \neq *$.

Cost analysis of the implementation

The following theorem shows that our implementation of the walk operator, as part of a space-efficient implementation of quantum backtracking, uses only a near-linear amount of qubits. Note that the routine implemented in this section have a polynomial time complexity.

Theorem C.1. *There is a polynomial-time implementation of the walk operator of the quantum backtracking framework for DPLL-like algorithms which uses at most $4n + w$ qubits, with $w \in O(\log(n))$.*

Proof. Having access to the unitary R_A and R_B, we implement the walk operator $R_B R_A$ with an ancilla qubit which checks whether the vertex $\vec{x}$

that we are considering is odd or even. Defining $\mathrm{Space}(U)$ to be the space (in terms of qubits) required by our implementation of a unitary U, we obtain that

$$\mathrm{Space}(R_B R_A) \leq \max(\mathrm{Space}(R_A), \mathrm{Space}(R_B)) + 1.$$

The operators R_A and R_B have very similar implementations, and under our implementation:

$$\begin{aligned}\mathrm{Space}(R_A) &\leq \mathrm{Space}(U_A) + 1 \\ &\leq \lceil \log_2(3) \rceil \cdot n + \mathrm{Space}(V_A) \\ &\quad + \mathrm{Space}(V_C) + O(1)\end{aligned}$$

where $\lceil \log_2(3) \rceil \cdot n$ corresponds to the space required to store a vertex, the implementation of V_A and V_C requires $\lceil \log_2(3) \rceil \cdot n + O(\log(n))$ additional ancillas (see Section C.4), adding up to an overall $4n + O(\log(n))$ space complexity for R_A (as $\lceil 2 \log_2(3) \rceil = 4$). □

C.4 Implementing the unit clause rule

In order to determine the next vertices in the search tree according to the unit rule, one needs to determine whether there exists a unit clause (i.e. a clause $C = \{l\}$ with only one literal l), given a partial assignment of the formula that we are considering. Whenever a unit clause is found, no branching occurs and the literal is simply set to true.

In order to implement this process, we need two operations:

- An operation $V_{\mathrm{unit}}^{(i)}$ which checks whether the i-th clause is a unit clause.
- An operation V_{next} which outputs the next partial assignment.

For each clause C_i, we implement the unitary

$$V_{\mathrm{unit}}^{(i)} : |\vec{x}\rangle|0\rangle_1|0\rangle_2 \mapsto |\vec{x}\rangle|j\rangle_1|s\rangle_2$$

which checks whether the i-th clause C_i is a unit clause under the partial assignment $\vec{x}$. It does so by checking whether the clause C_i is still 'alive' (not already satisfied under the partial assignment $\vec{x}$), and then checking whether it is a unit clause, to finally output whether it is a unit clause ($j \neq 0$), and if yes, the index j and polarity s of the forced variable. We let the unitary $\mathcal{C}_i$ determine whether C_i is alive by assignment $\vec{x}$ (by hardcoding the clause). And the unitary IsUnit_i checks whether $k-1$ variables of the clause C_i have been assigned a value: it is implemented with a counter from 0 till k, with an incrementation controlled on variables having a value different from $*$; and it outputs an index j and a sign s if C_i is a unit rule, 0 otherwise. Then the operation $V_{\mathrm{unit}}^{(i)}$, which checks whether the i-th clause C_i is a unit clause (and if yes output the index and sign of its variable), is implemented in Figure 7 ($\log(k)$-qubit counter ancilla is omitted).

In order to apply the unit rule, we apply $V_{\mathrm{unit}}^{(i)}$ to each clause C_i in the formula studied, and stop whenever we find a unit clause (i.e. whenever $V_{\mathrm{unit}}^{(i)}$ outputs $|\vec{x}\rangle|j\rangle|1\rangle$), see Figure 6 for the implementation of the unitary

$$V_{\mathrm{unit}} : |\vec{x}\rangle|0\rangle_3|0\rangle_4 \mapsto |\vec{x}\rangle|j\rangle_3|s\rangle_4$$

Note that, to go through all the L clauses of a formula, we need a clause counter which is implemented using $\lceil \log(L) \rceil \in O(\log(n))$ ancilla bits, since a k-SAT formula has at most $\binom{2n}{k} \in O(n^k)$ clauses.

If a unit clause is found, the current vertex only has one child, given by V_1, which is obtained by applying the following unitary V_{next} where b, j are the outputs of $V_{\mathrm{unit}}^{(i)}$:

$$V_{\mathrm{next}} : |\vec{x}\rangle|0\rangle|j\rangle|b\rangle \mapsto |\vec{x}\rangle|\vec{x}'\rangle|j\rangle|b\rangle$$

Such a unitary is implemented by copying $\vec{x}$ to the output register, while setting the j-th index to the value b.

If the unit rule is not applied, we can determine the j as next unassigned variable in partial assignment $\vec{x}$ (the first $*$), in a similar fashion as above. Then, we obtain V_1 (resp. V_2) by applying V_{next} with and $|b\rangle = |0\rangle$ (resp. $|b\rangle = 1$), so that given $|\vec{x}\rangle|0\rangle|j\rangle|0\rangle$, V_1 (resp. V_2) outputs $|\vec{x}\rangle|\vec{x}[x_j = 0]\rangle|j\rangle|0\rangle$ (resp. $|\vec{x}\rangle|\vec{x}[x_j = 1]\rangle|j\rangle|1\rangle$).

C.5 Implementing the pure literal rule

The pure literal rule eliminates variables x_i which only appear as the literal x_i or only appear as the literal $\overline{x_i}$. In which case, the variable is set to the value which makes the literal true, eliminating all the clauses which contains it in the process.

Classically, it is a convenient way to reduce the number of clauses manipulated by the backtracking algorithm considered. However, the quantum backtracking implementation that we present in

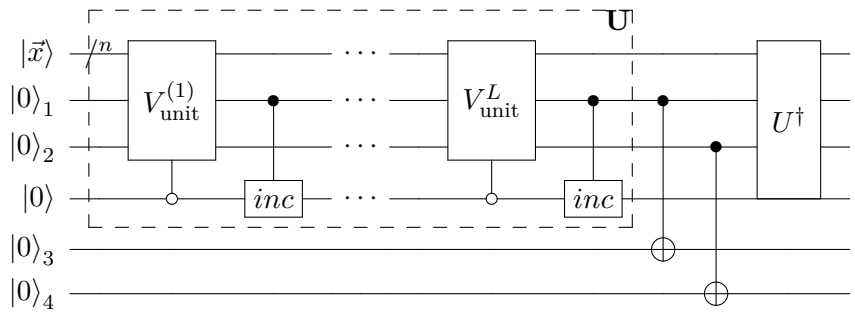

Figure 6: V_{unit} checks whether there is a unit clause

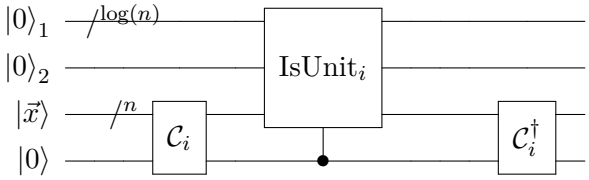

Figure 7: $V^{(i)}_{\text{unit}}$ checks whether C_i is a unit clause

this paper is not concerned with formula rewriting.

We implement a unitary

$$V_{\text{pure}} : |\vec{x}\rangle|0\rangle_1|0\rangle_2 \mapsto |\vec{x}\rangle|j\rangle_1|s\rangle_2$$

which checks whether there is a pure literal, and if yes, outputs want the index j of its variable and its polarity s. For completeness, $(j, s) = (0, 0)$ if no pure literal is found.

The implementation of the unitary V_{pure} is quite similar to the implementation of the unit rule. For each variable x_j, we only check whether it is a pure literal if no pure literal was found beforehand. In order to check whether a given variable x_j is part of a pure literal, we implement a unitary

$$V^{(i)}_{\text{pure}} : |\vec{x}\rangle|0\rangle_1|0\rangle_2 \mapsto |\vec{x}\rangle|j\rangle_1|s\rangle_2$$

which implements the following steps:

1. count the number of 'alive' clauses in which literals x_i and $\bar{x}_i$ respectively appear with two clause counters, by applying for each clause C_i the circuit $V^{(j,i)}_{\text{pure}}$ described in Figure 8 and uncomputing the ancillas in register 3 and 4;

2. use a Toffoli gate to determine whether only one of the counters is equal to 0, and if yes, we copy the index and sign of x_i in the output registers.

Then, the unitary V_{pure} simply applies $V^{(i)}_{\text{pure}}$ for each variable x_i, provided that no pure literal rule was found so far (this information is tracked with an ancilla variable counter of size $\lceil \log(n) \rceil$); and then uncompute the ancillas by applying the inverse of the circuit so far (as we did for the unit rule, see Figure 6).

This procedure is implemented using $O(\log(n))$ ancillas. As in the implementation of the unit rule (see Appendix C.4), we use V_{next} to compute the next partial assignment. Note that, we can implement in the same way a reduction rule which first applies the unit rule, and then applies the pure litteral rule if no unit clause is found.

C.6 Quantum tree size estimation

One drawback of Montanaro's quantum backtracking is that the runtime depends on the estimate of the size of the search tree (which is a parameter of the algorithm), and not on the size of the subtree that the classical backtracking algorithm explores.

The efficiency of a classical backtracking algorithm relies on its ability to explore the most promising branches first, which means than in practice, the algorithm may find a marked vertex after exploring T' vertices, where $T' \ll T$. Using a quantum tree size estimation subroutine to estimate the size of the subtree explored by the classical backtracking algorithm, the original quantum backtracking algorithm can be improved for the $T' \ll T$ case (see Th. C.2).

Theorem C.2 ([23])**.** *Consider a classical backtracking algorithm $\mathcal{A}$ which generates a search tree $\mathcal{T}$. There is a quantum algorithm which outputs 1 with high probability if $\mathcal{T}$ contains a marked vertex and 0 if it doesn't, with query complexity $O(n^{\frac{3}{2}}\sqrt{T'})$ where T' is the number of vertices actually explored by $\mathcal{A}$.*

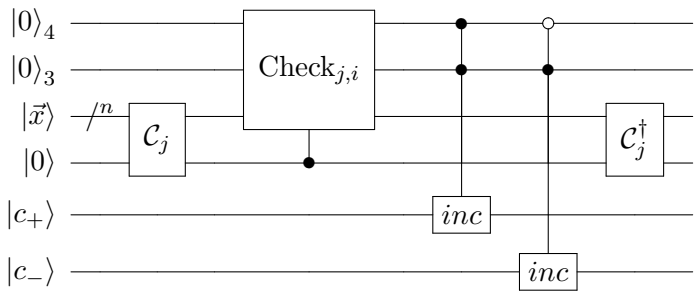

Figure 8: $V_{\text{pure}}^{(j,i)}$ checks whether the variable x_j appears in C_i and if yes, increases the counter corresponding to its polarity

The overall algorithm generates subtrees which contain the first 2^i vertices explored by the classical backtracking algorithm, increasing i until a marked vertex is found, or the whole search tree is considered. It is on each subtree containing the first 2^i vertices that we run the quantum backtracking algorithm.

Note that quantum backtracking with tree size estimation is only considered when $T' \ll T$. Because this algorithm is less performant than the original when T' is close to T, one can just switch to Montanaro's algorithm whenever the complexity of the generation of the path exceeds the complexity of the original quantum backtracking.

The main component of this variant of Montanaro's quantum backtracking are quantum backtracking itself and a quantum tree size estimation algorithm (Algorithm 1 in [23]), which both run QPE as a subroutine to detect whether the phase of a given unitary is equal to 0 (and therefore we can still apply the logarithmic space construction of Appendix B.2). Therefore it is sufficient to prove that quantum backtracking can be implemented efficiently in order to benefit from the speedup provided by Theorem C.2 in the hybrid framework.

D Eppstein's algorithm for the cubic Hamiltonian cycle problem

In [2], a hybrid divide-and-conquer algorithm for the Eppstein's algorithm for the cubic Hamiltonian cycle problem was provided based on Grover's search methods.

In the algorithm presented in [2], the quadratic improvement was achieved over the possible number of branching choices $(n/2)$ whereas the size of the overall search tree is upper bounded by $O(2^{n/3})$ (see [2, Appendix A] for detailed explanations[31]), where n represents the number of edges in the graph.

Consequently, Grover's approach yields a polynomial improvement in (it has a $O^*(2^{n/4})$ time complexity), which is not a near-quadratic speedup over classical $O^*(2^{n/3})$ time implementations of Eppstein algorithm (such quantum algorithm has an $O^*(2^{n/6})$ time complexity). This issue has been, outside of context of hybrid methods clarified and resolved in [25] using quantum backtracking.

In this section, we succinctly show how the quantum backtracking method for this problem can be applied in our hybrid context, achieving a near-quadratic speed up in the sub-tree, without any relevant loss off space efficiency (which translates to cut-points, and hence overall efficiency of the hybrid method) caused by relying on backtracking rather than Grover's search.

We first provide the following background for the benefit of the reader. The forced cubic Hamiltonian cycle problem (FCHC) is a NP-complete problem which asks whether a cubic graph $G = (V, E)$ (i.e., with degree 3) has a Hamiltonian cycle which contains at least all the edges in a given subset $F \subseteq E$.

Given a FCHC instance (G, F) as input, there is a classical divide-and-conquer algorithm $\mathcal{E}$ which solves the FCHC problem in time $O^*(2^{n/3})$ (a bound which constitutes an upper bound on the tree size) by selecting an unforced edge e and branching over the two subinstances $(G, F \cup \{e\})$ and $(G \setminus \{e\}, F)$ (see Algorithm 2 in [2]).

On a high-level, $\mathcal{E}$ is a backtracking algorithm

[31] This $n/2$ bound on the number of branching choices follows from the fact that the backtracking algorithm presented in [2] generates *full* binary trees of depth at most $s/2$ (see [2, Proposition 10]), where s is the effective problem size which is such that $s \leq n$.

which, on a given instance, does the following:

1. apply several reductions (in order to simplify the problem)
2. check whether one of the terminal conditions of the problem has been fulfilled (i.e. checks whether we can directly answer true or false)
3. choose the next edge to branch upon.

One can see $\mathcal{E}$ as an algorithm which explores binary search trees, whose root is labelled by the instance given as input, and each node (labelled by (G, F)) is either a leaf, or has two children (labelled by $(G, F \cup \{e\})$ and $(G \setminus \{e\}, F)$).

Now, for an implementation of quantum backtracking for algorithm $\mathcal{E}$ in the context of the hybrid approach, it is sufficient to implement space efficiently routines which checks whether the instance at any given node is a leaf (a routine denoted V_{leaf}), and otherwise what its children are (V_A), as explained in Appendix C.3.

In [2, Section V.C], reversible subroutines are introduced to re-construct the instance (labelling a node) from $\vec{v}$ (using $O(s\log(n/s) + s + \log(n))$ ancillas and $O(\mathsf{poly}(n))$ gates [2, Corollary 2]), and test whether it satisfies a terminal condition (using $O(\log(n))$ ancillas and $O(\mathsf{poly}(n))$ gates [2, Corollary 3]), where s is the effective problem sized defined by $s = n - |F| - |C(G, F)|$ and $C(G, F)$ is the set of 4-cycles which are disconnected from F.

Now, we take those subroutines as an implementation V_{leaf}, the routine of quantum backtracking which checks whether the current node is a leaf. The algorithm $\mathcal{E}$ always branches over two choices (adding or removing an edge). In the implementation of quantum backtracking for $\mathcal{E}$, one can construct an unitary V_A which given the label $\vec{v} = v_1 \dots v_i$ of the current node, determines the label of its children $\vec{w}$ and $\vec{w}'$, which are respectively $v_1 \dots v_i 0$ and $v_1 \dots v_i 1$ (see Appendix C.3).

E Reversible Simulation of SIA

This appendix implements the circuit SIA of Section 4.2 reversibly for formulas of bounded-index width (biw; see Definition 4.2) in polynomial time and in a space-efficient way (as a function of w and n). We assume the input is a formula F defined on n variables with biw w.

For space efficiency, as discussed in Section 4.2, we will not manipulate the formula F, but instead work on partial assignments in the branching representation, which we call the *advice* (see Section 2.4.1). The downside of this approach is of course that for a given node label, our circuits need to reconstruct the corresponding partial assignment. To avoid using $n\log(3)$ qubits for reconstructing the partial assignment, we exploit the bounded index width to split the computation into blocks of w variables. We then use Bennett's pebbling strategy [29] to compose these blocks into a space-efficient reversible circuit which computes SIA.

The following three subsections treat the reconstruction of the partial assignments (Sec. E.1), split the computation into blocks of w variables (Sec E.2), and combine these subcomputations in a space-efficient way (Sec E.3). At the end of this we obtain Theorem E.1.

Theorem E.1. *There is a reversible circuit that, given an advice of size S_{adv}, computes s-SIA for a w-biw formula F with n variables in space*

$$S_w \triangleq O(w \cdot \log(n/w))$$

and time

$$O((n/w)^{\log(3)} \cdot w \cdot (2+sw)^{ks} \cdot \mathsf{polylog}(n))$$

Note that the space complexity S_w excludes the space required to store the advice (S_{adv}). The space requirements S_w and S_{adv} are also treated separately in Section 4.2. As a final remark we also note that while in practice the size of s for s-implication can be considered a large constant, it should in fact grow (slowly) in n [5] (the inverse Ackermann function would suffice).

E.1 Reconstructing partial assignments

The simplifications in the search algorithm discussed in Section 4.2 (see Algorithm 3) shift the difficulty to creating a space-efficient unitary implementing the search predicate P. It should take as input an advice string $\vec{a}$ of length $S_{adv} = |\vec{a}|$, representing the node in the search tree and determine whether this node is a solution leaf ($P(\vec{a}) = 1$), a normal leaf ($P(\vec{a}) = 0$) or a branch (such that we can query children $ch2(\vec{a}, 0)$ and $ch2(\vec{a}, 1)$). For conciseness, we will assume here that the advice length S_{adv} is fixed and no counter is required for recording the length of $\vec{a}$.

Algorithm 5 SIA specification for reconstructing the full partial assignment from the advice in branching representation (indices start at 1).

```
SIA_F (a⃗)
    p := 1                                  ▷ advice pointer
    for i in 1 .. n
        if F_{x_1,..,x_{i-1}} = b
            return b                        ▷ 0 children
        if F_{x_1,..,x_{i-1}} ⊨_s x_i ∨ F_{x_1,..,x_{i-1}} ⊨_s x̄_i
            x_i := F_{x_1,..,x_{i-1}} ⊨_s x_i
        else if p = |a⃗|                    ▷ out of guesses
            return ⊥                        ▷ 2 children
        else
            x_i := a⃗[p]
            p := p + 1
```

Ignoring space requirements for a moment, Algorithm 5 provides a specification of $P(\vec{a})$. We call this procedure s-implication with advice (SIA). It reconstructs the partial assignment on variables $x_1, \ldots, x_n$ by testing s-implication on each variable x_i in order. If x_i is forced (s-implied) the variable is assigned accordingly. If x_i should be guessed, then the next value from the advice is used to assign it, using the advice pointer p. If at one point the formula becomes trivial under the partial assignment, i.e., $F_{|\vec{x}} = b$, for $b \in \{0, 1\}$, then the specification returns b (simplifying from the $|1b\rangle$ discussed in Section 4.2). If the advice runs out ($p = |\vec{a}|$) and the formula is not yet trivial after eagerly forcing variables then the return value is $\perp$ ($|00\rangle$).

E.2 Splitting SIA into blocks

In order to create a space-efficient, reversible implementation of the entire SIA operation, we split SIA into a number of blocks, each of which acts on a subset of variables. The variables $x_1, \ldots, x_n$ of F are split up into $l = n/w$ blocks of size w (for simplicity we assume w divides n). Each block B_i, for $i \in \{1, \ldots, l\}$, contains variables

$$B_i \triangleq \{x_{j+1}, \ldots, x_{j+w-1}\}, \text{ with } j = (i-1)w. \tag{14}$$

Note that having a bounded index width of w guarantees that computing the s-implication of x_i can only rely on $x_{i-w}, \ldots, x_{i-1}$, and so in order to compute SIA for variables in B_i we only need the assignments on the previous w-sized block B_{i-1}. This is made more explicit in Appendix F.1.

We define a subcircuit $\texttt{SIAB}_i$ (see Figure 9) which computes SIA (Algorithm 5) for variables in B_i, unless input r (a flag) is set, in which case it is the identity function. Moreover, if $\texttt{SIAB}_i$ encounters a contradiction or runs out of advice, it sets r. The first three registers contain the inputs: an assignment $\vec{y}_{i-1}$ on the variables in B_{i-1}, the advice pointer p (of $\log(n)$ qubits) which keeps track of the next unused advice bit, and the flag r (1 qubit). The outputs are written to the next three registers and contain an assignment $\vec{y}_i$ to the variables in B_i, and the updated values of p and r. Additionally, each $\texttt{SIAB}_i$ block uses ancilla qubits $\vec{a}_b$. These ancilla's are initialized in an all-zero state, and are reset to zero by means of uncomputation within the $\texttt{SIAB}_i$ block. As such, this ancilla register can be reused between all the blocks. Finally, $\texttt{SIAB}_i$ needs read-only access to the advice $\vec{a}$, which is also shared between all $\texttt{SIAB}_i$ blocks. Before $\texttt{SIAB}_1$ the registers for p, r, and $\vec{a}_b$ are initialized to all zeros.

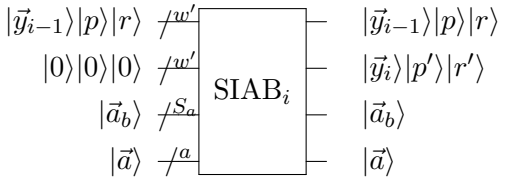

Figure 9: A $\texttt{SIAB}_i$ block, for $i \geq 2$. $\texttt{SIAB}_1$ is a simpler version of $\texttt{SIAB}_i$ without the top register. Here $w' = w + \log(n) + 1$, $S_a = w + O(\log(n))$, and $a = S_{adv}$.

A more detailed implementation of $\texttt{SIAB}_i$ is given in Appendix F.

E.3 Combining SIAB blocks

Having defined $\texttt{SIAB}_i$, the composition $\texttt{SIAB}_l \circ \ldots \circ \texttt{SIAB}_1$ now computes the value we are interested in, still disregarding the low-space requirement. Note that in order to "compose" two blocks $\texttt{SIAB}_{i+1} \circ \texttt{SIAB}_i$, the second output register (Fig. 9) of $\texttt{SIAB}_i$ needs to be connected to the first input register of $\texttt{SIAB}_i$. A circuit which composes the $\texttt{SIAB}_i$ blocks in this manner is shown in Figure 10. In order to keep to computation unitary, the top output register of each $\texttt{SIAB}_i$ block cannot simply be discarded, even though these registers are not used as inputs anymore.

Remark E.2. *Naively composing all $\texttt{SIAB}_i$ by simply allocating a new w-sized register for each block (see Fig. 10) requires $l \cdot w' + S_a$ qubits on top of the S_{adv} qubits storing the advice.*

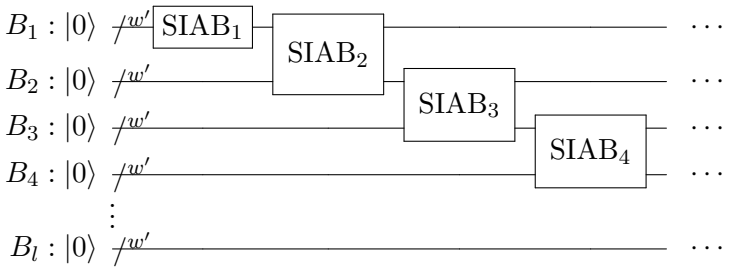

Figure 10: A (partial) visualization of the naive composition of all $\mathtt{SIAB}_i$ blocks. The bottom register of each $\mathtt{SIAB}_i$ block (see Fig. 9) has been omitted from this figure. This method requires $l \cdot w' + S_a$ qubits

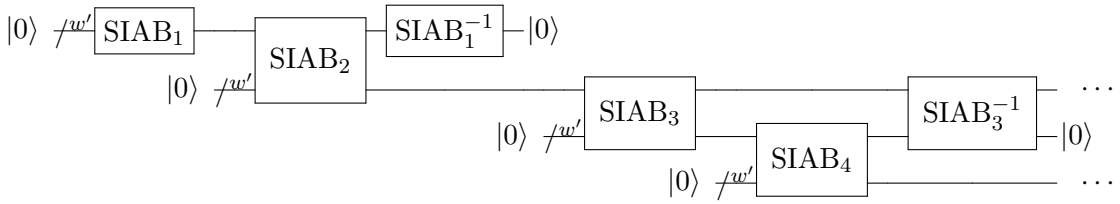

Figure 11: A (partial) visualization of $\mathtt{SIAR}$; the composition of the $\mathtt{SIAB}_i$ blocks using Bennett's algorithm [29]. The bottom register of each $\mathtt{SIAB}_i$ block (see Fig. 9) has been omitted from this figure. Uncomputing $\mathtt{SIAB}_1$ frees up a w' sized register which can be used by $\mathtt{SIAB}_3$. This method requires $w' \cdot (\log(l) + 1) + S_a$ qubits.

To meet the space requirement of $o(n)$, we must reduce the $l \cdot w = n$ qubits used by combining block. To achieve this, we turn to Bennett's strategy [29]. It relies on reducing the problem of turning a deterministic computation into a space-efficient reversible computations to a *reversible pebble game*. Such a pebble game is played on a rooted, directed graph, which corresponds to the dependency graph of inputs and outputs of different sub-computations. The rules of a reversible pebble game can be defined as follows: one can (un)pebbled a leaf freely, whereas any other node can only be (un)pebbled if their predecessors are pebbled; one wins the game if one can put a pebble on the root node. After the game is won, the maximum number of pebbles used at any time during the game corresponds to the amount of space the reversible circuit requires.

For our problem (composing $\mathtt{SIAB}_l \circ ... \circ \mathtt{SIAB}_1$) the dependency graph is simply a line graph of length l. Bennett's algorithm constructs a concrete pebbling strategy for the line graph corresponding to a deterministic computation. The strategy shows that a line graph of length l can be reversibly pebbled with only $\lfloor \log(l) \rfloor + 1$ pebbles. A clear visualization of this algorithm is shown in [29, Table 2]. Figure 11 shows the $\mathtt{SIAB}_i$ blocks composed using Bennett's algorithm, and illustrates how this saves on the space requirements: For example, by temporarily uncomputing $\mathtt{SIAB}_1$, a w' sized register is freed up which can be used by $\mathtt{SIAB}_3$. We will refer to this specific reversible composition of the $\mathtt{SIAB}_i$ blocks as $\mathtt{SIAR}$.

Lemma E.3 (Rephrased from Bennett [29]:)**.** *The composition of (irreversible) operations $f_t \circ \cdots \circ f_1$, each taking T time and S space, can be implemented reversibly in $O(t^{\log(3)} \cdot T)$ time and $(\lfloor \log(t) \rfloor + 1) \cdot S$ space.*

Theorem E.4. *There is a reversible circuit, $\mathtt{SIAR}$, that computes SIA for a w-biw formula F using only*

$$T_r \triangleq O(l^{\log(3)} \cdot w \cdot (2 + sw)^{ks} \cdot \mathsf{polylog}(n))$$

time and

$$S_r \triangleq (\lfloor \log(l) \rfloor + 1) \cdot w' + S_a = O(w \cdot \log(l))$$

space (wires), where $w' = (w + \log(n) + 1)$ and $S_a = w + O(\log(n))$.

Proof. By Theorem F.1, $\mathtt{SIAB}$ can be reversibly implemented in $T_b \triangleq O(w \cdot (2+sw)^{ks} \cdot \mathsf{polylog}(n))$ time. The time for $\mathtt{SIAR}$ follows then from Lemma E.3. In terms of space, all $\mathtt{SIAB}$ blocks share a $S_a = O(w + \log(n))$ sized register, on top of which Bennet's strategy requires $(\lfloor \log(l) + 1 \rfloor) \cdot w'$ space to reversibly compose all $\mathtt{SIAB}_i$ (Lem. E.3). □

Theorem E.1 simplifies this statement.

F Implementation of SIAB

In this note, we implement the circuit $\texttt{SIAB}_i$ defined in Appendix E, which realizes a w-sized block of computation of s-SIA, by computing the values of the variables in B_i (Eq. 14) given access to the variables in B_{i-1}. First, the dependency of $\texttt{SIAB}_i$ on only the previous w assigned variables is proven in Section F.1. After that, the specification of the circuit $\texttt{SIAB}_i$ (given in Sections F.2 and F.3) ultimately leads to a proof of the space and time complexity (see Sec. F.4) stated in following theorem.

Theorem F.1. *Each circuit $\texttt{SIAB}_i$ can be implemented reversibly in $T_b \triangleq O(w \cdot (2+sw)^{ks} \cdot \mathsf{polylog}(n))$ time, with $S_b \triangleq 2(w+\log(n)+1)$ wires and $S_a \triangleq w + O(\log(n))$ ancillas, omitting the space for the advice $\vec{a}$.*

F.1 S-implication under bounded index width

In this subsection we prove that in order to do s-implication in the dncPPSZ algorithm for a variable x_j, only two sets of variables need to be considered: the previous w assigned variables, and any of the unassigned variables x_k, with $j \le k \le n$.

We recall the definition of s-implications.

Definition F.2. *A literal x_j ($\overline{x_j}$) is s-implied by F, written $F \models_s x_j$ ($F \models_s \overline{x_j}$), if and only if there is a subset of clauses $G \subseteq F$ of size at most s such that all satisfying assignments of G set x_j to 1 (0).*

The following lemma specifies that it is not necessary to remember a partial assignment in its entirety in order to determine s-implications.

Lemma F.3. *Consider an n-variables formula F of bounded index width w, with a fixed variables order $x_1 < \ldots < x_n$. Let $F_{|\alpha}$ be the formula obtained by assigning $\alpha = [\alpha_1, \ldots, \alpha_{j-1}]$ to variables $[x_1, \ldots, x_{j-1}]$, and assume that $F_{|\alpha}$ is not yet trivial. Then, in order to determine whether $F_{|\alpha} \models_s x_j$, it is sufficient to consider the assignment $\beta = [\alpha_{j-w}, \ldots, \alpha_{j-1}]$ to $[x_{j-w}, \ldots, x_{j-1}]$.*

Proof. Let us define F_1, F_2, and F_3 as subformulas of F, such that each only contains clauses with the variables as visualized in Figure 12. By the property of index width, we have that the formula $F_{|\beta}$ consists of two independent formulae F_1 and F_3. Hence, we have $F_3 = F_{2|\beta} = F_{|\alpha}$. It immediately follows then that $F_{|\alpha} \models_s x_j \iff F_{2|\beta} \models_s x_j$. □

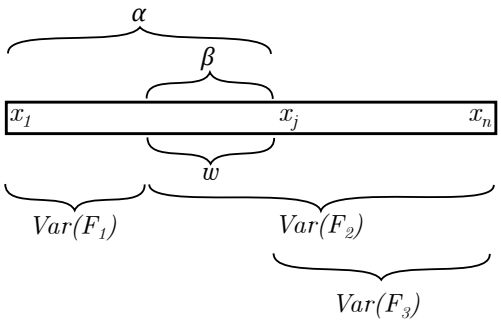

Figure 12: Visualization of subsets of variables of F.

Generally, in order to check if $F_{|\alpha} \models_s x_j$, all s-sized $G \subseteq F$ need to be considered, where each G can take variables from $\{x_1, \ldots, x_n\}$. However, since F has bounded index width w, from Lemma F.3 we have that $F_{2|\beta} = F_{|\alpha}$ and so checking all $G \subseteq F_2$ is sufficient.

F.2 Implementing $\texttt{SIAB}_i$

Algorithm 6 gives a high level presentation of the implementation of $\texttt{SIAB}_i$, with references to the corresponding reversible subroutines (see Section F.3). We omit $\texttt{SIAB}_1$ which is a simpler version of $\texttt{SIAB}_i$ without input block. We write $\leftarrow$ for the action of copying the value of a bit with a XOR gate.

Algorithm 6 Pseudocode for $\texttt{SIAB}_i(\vec{y}_{i-1}, p, r)$.

```
for j in iw ... iw + w − 1                          ▷ Fig. 15
  if (r = 0)
    c ← 0
    for each s-sized G ⊆ F_{2|ȳ_{i−1}}              ▷ Fig. 14
      if (c = 0)
        (b′, b′_j) ← check if G ⊨ x_j, x̄_j          ▷ Fig. 13
        if (b′ = 1)                                  ▷ x_j is implied
            c ← c ⊕ 1
        if G is unsat
            c ← c ⊕ 1
            r ← r ⊕ 1
    if (b′ = 1)                                      ▷ x_j is implied
        x_j ← b′_j
    else                                             ▷ A in Fig. 14
        if p < a
            x_j ← a⃗[p]
            p ← p + 1
        else
            r ← r ⊕ 1
o1, o2, o3 ← [x_{i·w}, ..., x_{(i+1)·w}], p, r
uncompute workspace                                  ▷ L† in Fig 15
```

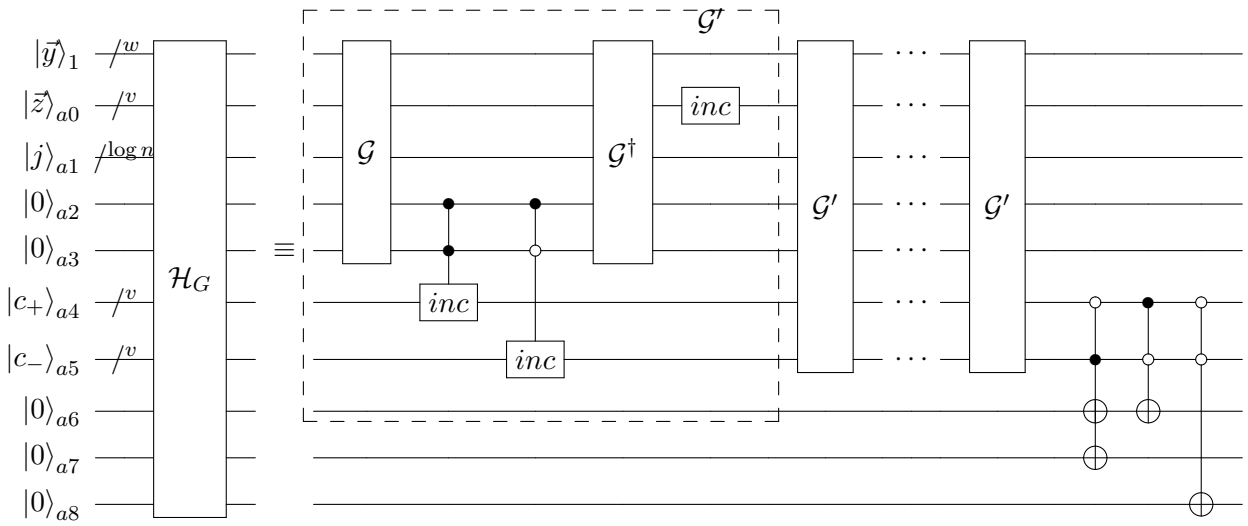

Figure 13: Routine $\mathcal{H}_G$, used to check whether x_j is s-implied by a given subset of clauses G. Each $\mathcal{G}'$ checks a single assignment $\vec{y}$ to the variables in G. The $\mathcal{G}'$ block is repeated $|V'|$ times. The registers for $\vec{z}$, c_+, and c_- need $v = \lceil \log(|V'|) \rceil \leq ks$ qubits. The controls on $|c_+\rangle_{a4}$ and $|c_-\rangle_{a5}$ check whether $c_\pm = 0$ (open dot) or $c_\pm \neq 0$ (solid dot).

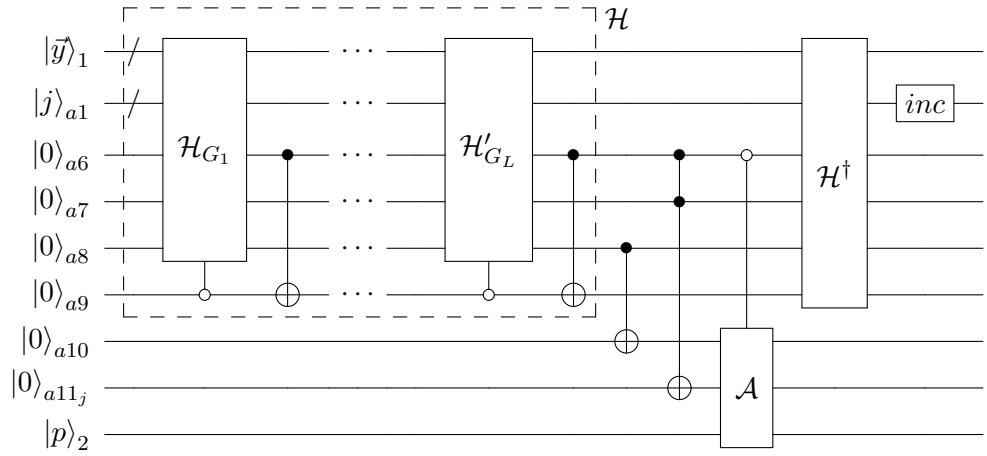

Figure 14: Inner loop $\mathcal{L}_j$ of $\texttt{SIAB}_i$, checking all s-sized subsets of clauses and fixing the value of the current variable. For ease of visualization, wires in $\mathcal{H}_G$ (Fig. 13) which are not directly relevant in $\mathcal{L}_j$ have been omitted.

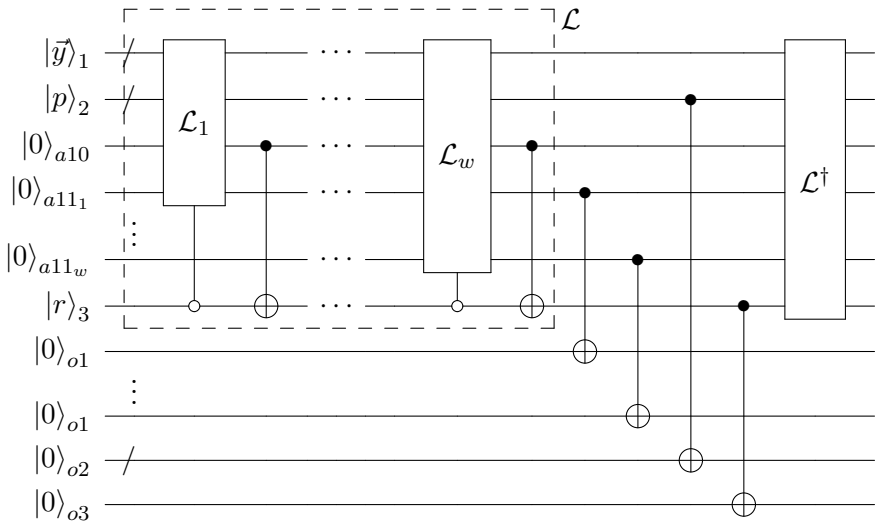

Figure 15: Circuit for $\texttt{SIAB}_i$, running the inner loop for every variable in the current w-block, provided that the remaining formula has not been found to be unsatisfiable (unsatisfiability would be indicated by $r > 0$). For ease of visualization, wires in $\mathcal{L}_j$ (Fig. 14) which are not directly relevant this high-level overview have been omitted.

Each circuit $\texttt{SIAB}_i$ has the following input registers: register 1 contains an assignment $\vec{y}_{i-1}$ to previous w-sized block B_{i-1}, register 2 contains the pointer p to the next unused advice index (defined as a counter from 0 to $S_{adv}-1 = |\vec{a}|-1$, using $\lceil \log(S_{adv}) \rceil$ bits), register 3 contains a one bit flag r which is raised whenever we encounter a contradiction or the advice is fully used. The registers $o1, o2, o3$ are used to output the next values $\vec{y}_i$, p', and r' respectively. We also allocate ancilla registers $a0, \ldots, a11$.

$$\texttt{SIAB}_i\colon |\vec{y}_{i-1}\rangle_1|p\rangle_2|r\rangle_3|0\rangle_{o1}|0\rangle_{o2}|0\rangle_{o3}|\vec{a}\rangle|\vec{a}_b\rangle \mapsto |\vec{y}_{i-1}\rangle_1|p\rangle_2|r\rangle_3|y_i\rangle_{o1}|p'\rangle_{o2}|r'\rangle_{o3}|\vec{a}\rangle|\vec{a}_b\rangle$$

The ancilla registers $a0, \ldots, a8$ are the ones primarily manipulated by the subroutines $\mathcal{H}$ and $\mathcal{H}_G$, as defined in Section F.3. Ancilla registers $a9$ and $a10$ are used to store c' and r' (see Alg. 6), such that c and r can be modified within the loop-body, without immediately affecting the control qubits c' and r'. Finally, ancilla register $a11$ is a buffer of size w in which one stores the values of the current w-block B_i.

The circuit in Figure 14 corresponds to the the inner loop of $\texttt{SIAB}_i$ and the subsequent instructions which fix the value of x_j. The unitary $\mathcal{A}$ corresponds to the instructions in the pseudocode which set x_j to the next advice bit and increase the advice pointer. Figure 15 gives the implementation of the whole circuit $\texttt{SIAB}_i$ (note that running $\mathcal{L}_j$ is conditional on $r'=0$, and the circuit flips r when it finds that the remainder of the formula is unsatisfiable). To simplify our representation, we omit ancilla wires which are only manipulated in the subcomponents of the circuit, but not the overall circuit.

F.3 Reversible subroutines of $\texttt{SIAB}_i$

For each s-sized subformula $G \subseteq F$, we define a unitary

$$\mathcal{G}: |\vec{y}\rangle_1|\vec{z}\rangle_{a0}|j\rangle_{a1}|0\rangle_{a2}|0\rangle_{a3} \mapsto |\vec{y}\rangle_1|\vec{z}\rangle_{a0}|j\rangle_{a1}|b\rangle_{a2}|b_j\rangle_{a3}$$

which takes the following inputs: an index j, an assignment $\vec{y}$ to the w variables preceding x_j, an assignment $\vec{z}$ of length (at most) ks to variables of indices greater or equal to j, and two ancilla qubits to write the output to.

Each unitary $\mathcal{G}$ hardcodes the Boolean formula G, and encodes the following reversible procedure: if variable x_j does not appear in G, output $(0,0)$; else if G is satisfied by the partial assignment specified by $\vec{y}$ on $x_{j-w}, \ldots, x_{j-1}$ and $\vec{z}$ on

$$V' = \mathrm{Var}(G) \setminus \{x_{j-w}, \ldots, x_{j-1}\},$$

output $(1, b_j)$, where b_j is the assignment to variable x_j, and else output $(0,0)$.

Next, we define a unitary

$$\mathcal{G}': |\vec{y}\rangle_1|\vec{z}\rangle_{a0}|j\rangle_{a1}|c_+\rangle_{a4}|c_-\rangle_{a5} \mapsto |\vec{y}\rangle_1|\vec{z}\oplus 1\rangle_{a0}|j\rangle_{a1}|c'_+\rangle_{a4}|c'_-\rangle_{a5}$$

which applies $\mathcal{G}$, increase the counter c_+ (c_-) if $\mathcal{G}$ outputs $(1,1)$ ($(1,0)$), and then uncomputes the ancilla registers. Additionally $\mathcal{G}'$ increased the value of the $|\vec{z}\rangle_{a0}$ register, such that the next $\mathcal{G}'$ is applied to the next partial assignment. This is shown in Figure 13. The registers $|c_+\rangle_{a4}$ and $|c_-\rangle_{a5}$ (as well as $|\vec{z}\rangle_{a0}$) require $\lceil \log(|V'|) \rceil \leq ks$ qubits to count up to $2^{|V'|}$, the total number of possible assignments to G.

For each subformula $G \subseteq F$, we implement an unitary

$$\mathcal{H}_G: |\vec{x}\rangle_1|j\rangle_{a1}|0\rangle_{a6}|0\rangle_{a7}|0\rangle_{a8} \mapsto |\vec{x}\rangle_1|j\rangle_{a1}|b'\rangle_{a6}|b'_j\rangle_{a7}|r''\rangle_{a8}$$

which evaluates G for all its possible partial assignments, to determine whether the satisfying assignments of G agree on the value assigned to x_j, and if yes ($b'=1$), outputs that value (b'_j). If G is unsatisfiable, $\mathcal{H}_G$ outputs $r''=1$.

Our implementation of $\mathcal{H}_G$ initializes the ancilla registers $a0, a2, \ldots, a5$ to 0, and repeatedly applies $\mathcal{G}'$ and increments $\vec{y}$ (see Figure 13) up until $\vec{y} = 2^{|V'|}-1$. After applying $\mathcal{G}'$ for all assignments $\vec{y}$, the following can be concluded based on the counters c_+ and c_-:

$$\begin{cases} x_j \text{ not implied, } G \text{ is sat} & \text{if } c_+ > 0, c_- > 0 \\ G \text{ implies } x_j & \text{if } c_+ > 0, c_- = 0 \\ G \text{ implies } \overline{x_j} & \text{if } c_+ = 0, c_- > 0 \\ G \text{ is unsat} & \text{if } c_+ = 0, c_- = 0 \end{cases}$$

If both c_+ and c_- are non-zero, G has satisfying assignments, but they do not all agree on x_j. In this case $\mathcal{H}_G$ outputs $(b', b'_j) = (0,0)$. If exactly one of c_+ and c_- is non-zero, all satisfying assignments of G agree on x_j, and either x_j or $\overline{x_j}$ is implied by G. Here $\mathcal{H}_G$ outputs $(1,1)$ or $(1,0)$ respectively. If both c_+ and c_- are zero, G is not satisfiable at all given the variables assigned before x_j, and so neither is F. In this case $\mathcal{H}_G$ outputs $r''=1$, which ultimately leads the circuit to raise the flag r.

F.4 Complexity of the implementation of $\texttt{SIAB}_i$

First, let us analyze the time complexity of $\texttt{SIAB}_i$. To simplify the analysis, we drop any $\mathsf{polylog}(n)$ factor by writing $\tilde{O}(f(n)) = O(\mathsf{polylog}(n) \cdot f(n))$. By doing so the increment operations, as well as the $\mathcal{A}$ operation can be done in $\tilde{O}(1)$ time.

The unitary $\mathcal{G}$ hardcodes a Boolean formula G which has (at most) s clauses and k variables per clause, taking $T_{\mathcal{G}} = O(ks) = \tilde{O}(1)$ time. The subroutine $\mathcal{G}'$ runs $\mathcal{G}$ and $\mathcal{G}^\dagger$, and a number of $\tilde{O}(1)$ operations, which gives $T_{\mathcal{G}'} = \tilde{O}(1)$. Next, $\mathcal{H}_G$ repeats $\mathcal{G}'$ for all assignments to $\vec{z}$, which are at most 2^{ks}, giving $T_{\mathcal{H}_G} = \tilde{O}(2^{ks})$.

The inner loop $\mathcal{L}_i$ loops over all s-sized formulae G. Since we consider formulas with bounded index width w, for any variable x_j, any subformula $G \subseteq F$ with s clauses can either be split into independent formulae or contains only unassigned variables from $\{x_j, \dots, x_{j+sw}\}$ (a chain of s clauses each with index width w). Each s-sized G has at most ks variables, and which sw options per variable this results in at most $O((sw)^{ks})$ formulae G. This yields a time complexity of $T_{\mathcal{L}_j} = \tilde{O}((sw)^{ks} \cdot 2^{ks}) = \tilde{O}((2+sw)^{ks})$.

Finally, the outer loop of $\texttt{SIAB}_i$ loops over w variables. Giving $T_b = \tilde{O}(w \cdot (2+sw)^{ks}) = O(w \cdot (2+sw)^{ks} \cdot \mathsf{polylog}(n))$.

Including the input registers $1, 2, 3$ and output registers $o1, o2, o3$ the space requirement for is $\texttt{SIAB}_i$ is $2(w + \log(n) + 1) + S_a$, where S_a is the space for the ancilla registers $a0$ through $a11$, which adds up to $S_a = w + \log(n) + 3ks + O(1) = w + O(\log(n))$.

G Lattice SAT is NP-complete

We define Lattice SAT to be the restriction of the 3-SAT problem to formulas defined on a lattice, so that each clause is associated to a constraint defined on a tile (unit square of the grid), with corners of the tiles labelled by variables.

Formally, an instance of Lattice SAT is a formula defined on n variables, and whose clauses are defined by 2 to 3 corners on the same tile (see Figure 16), with at least one tile defined on 3 corners, in such a way that the overall lattice fits within a square of length $\sqrt{n}$.

By construction, there are permutations of indices which are such that any given instance of Lattice SAT has bounded index width $\sqrt{n} \in o(n)$.

First, observe that Lattice SAT is in NPbecause the validity of an assignment for the underlying 3-SAT instance can be verified in polynomial time. Let us show that Lattice SAT is NP-hard.

A polynomial-time reduction from 3-SAT to Lattice SAT can be implemented as follows. Consider a 3-SAT formula F defined on n variables over L clauses. Consider a lattice $\mathbb{F}$ defined as a nL-by-nL 2-dimensional grid. In what follows, for each clause C, we place variables (which appear in C) in the lattice $\mathbb{F}$, and we add to the lattice $\mathbb{F}$ a set of tiles which is equisatisfiable to the clause C.

We associate L copies $x_{i,1}, \dots, x_{i,L}$ (one per clause) to each variable x_i. We introduce new constraints to ensure that all copies have the same truth value, that is

$$x_{i,l} \vee \bar{x}_{i,l+1} \text{ and } \bar{x}_{i,l} \vee x_{i,l+1}.$$

Having L copies of each variable x_i allows for a spatial arrangement on the lattice of any two copies $x_{i,l}$ and $x_{i,l+1}$ in a constrained relationship, by defining one tile in the lattice where they meet. We place such copies on the diagonal of the lattice $\mathbb{F}$, from the top left corner to the bottom right one in the order determined by the order of variables in $\mathrm{Vars}(F)$.

Consider a clause $C_l = x_{i_1} \vee x_{i_2} \vee x_{i_3}$ in F (where we assume without loss of generality that i_1, i_2 and i_3 are variables indices such that $i_1 \leq i_2 \leq i_3$). Let us define a set of tiles which corresponds to a restricted formula which is equisatisfiable to C_l. Observe that each clause C_l can be decomposed into three clauses $C'_l = x_{i_1,l} \vee x_{i_2,l} \vee t$, $C''_l = x_{i_2,l} \vee x_{i_3,l} \vee \bar{t}'$ and $C'''_l = \bar{t} \vee t'$ (for some fresh variables t, t'), so that $C_l \equiv C'_l \wedge C''_l \wedge C'''_l$.

In what follows, we construct a set of constraints which is equisatisfiable to the clause C'_l. First, fresh variables $y_1, \dots, y_p, z_1, \dots, z_q, t$ are spatially arranged on the lattice $\mathbb{F}$ so that: $y_1, \dots, y_p$ are placed on the same horizontal line as $x_{i_1,l}$ in $\mathbb{F}$, with y_1 directly at the right of $x_{i_1,l}$, and each y_{i+1} is placed directly at the right of y_i; $z_1, \dots, z_q$ are placed on the same vertical line as $x_{i_2,l}$, with z_1 directly above $x_{i_2,l}$, and each z_{i+1} is placed directly above z_i; t is placed on the same tile as y_p and z_q and t, so that t is directly at the right of y_p and directly above z_q.

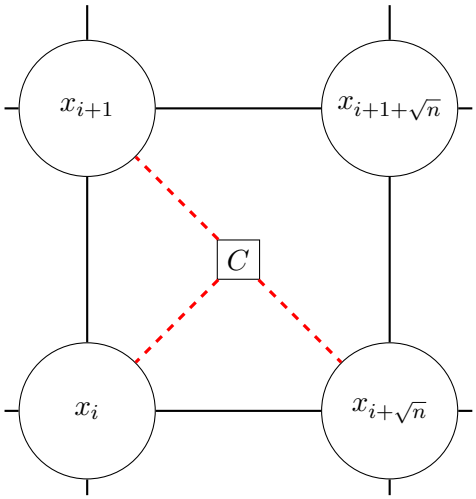

Figure 16: $C = x_i \vee \neg x_{i+\sqrt{n}} \vee x_{i+1}$ in Lattice SAT (black lines) and in Planar 3-SAT (red dotted lines)

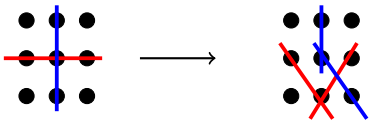

Figure 17: Rewriting overlaps

Then, we add the following set of the constraints to the tiles of lattice $\mathbb{F}$ on which those variables $y_1, \ldots, y_p, z_1, \ldots, z_q, t$ are located:

- $\bar{y}_i \vee y_{i+1}$, $\bar{y}_{i+1} \vee y_i$ (for $0 \leq i \leq p$),
- $\bar{z}_j \vee z_{j+1}$, $\bar{z}_{j+1} \vee z_j$ (for $0 \leq j \leq q$),
- $y_p \vee t \vee y_1$,

where y_0 is $x_{i_1,l}$ and z_0 is $x_{i_2,l}$, ensuring that the variables $y_1, \ldots, y_p, x_{i_1,l}$ take the same truth value, and the variables $z_1, \ldots, z_q, x_{i_2,l}$ take the same truth value.

We repeat the same process for x_{i_2} and x_{i_3}, adding fresh variables $y'_1, \ldots, y'_p, z'_1, \ldots, z'_q, t'$, with the same constraints but this time $\bar{t}'$ in the constraint where t previously appeared. We obtain a second set of constraints which is equisatisfiable to C''_l.

Now, observe that reiterating the same process a third time for t and t', we obtain a third set of constraints which is equisatisfiable to C'''_l. Combining the three sets of constraints, we obtain a set of constraints which is equisatisfiable to the formula $C'_l \wedge C''_l \wedge C'''_l$, which is itself equisatisfiable to the clause C_l.

We repeat this process for every clause C_l, obtaining an instance $\mathbb{F}$ of Lattice SAT defined by $O((nL)^2)$ constraints on a nL-by-nL squared grid. The instance $\mathbb{F}$ is not equisatisfiable to F, as our reduction potentially generate overlaps, which can be eliminated by repeatedly inserting empty rows and lines and applying the rewriting gadget described in a simplified representation in Figure 17, with coloured lines corresponding to tiles defined on two variables (black dots). At most $(nL)^2$ overlaps exist, and $L \in O(\text{poly}(n))$, so that this reduction can be done in polynomial time.

Theorem G.1. *Lattice SAT is* NP*-complete.*